Applications and Industry®

July 1959

Transactions Papers

NUMBER 43

Published Bimonthly by

AMERICAN INSTITUTE OF ELECTRICAL ENGINEERS

Communication and Electronics—July 1959

(See inside back cover)

Note to Librarians. The six bimonthly issues of "Applications and Industry," March 1959–January 1960, will also be available in a single volume (no. 78) entitled "AIEE Transactions—Part II. Applications and Industry," which includes all technical papers on that subject presented during 1959. Bibliographic references to Applications and Industry and to Part II of the Transactions are therefore equivalent.

Applications and Industry. Published bimonthly by the American Institute of Electrical Engineers, from 20th and Northampton Streets, Easton, Pa. AIEE Headquarters: 33 West 39th Street, New York 18, N. Y. Address changes must be received at AIEE Headquarters by the first of the month to be effective with the succeeding issue. Copies undelivered because of incorrect address cannot be replaced without charge. Editorial and Advertising offices: 33 West 39th Street, New York 18, N. Y. Nonmember subscription $8.00 per year (plus 50 cents extra for foreign postage payable in advance in New York exchange). Member subscriptions: one subscription at $2.50 per year (balance of $5.00 subscription price to be paid by application of annual dues) to any one of three divisional publications: Communication and Electronics, Applications and Industry, or Power Apparatus and Systems: additional annual subscriptions $5.00 each. Single copies when available $1.50 each. Second-class mail privileges authorized at Easton, Pa. This publication is authorized to be mailed at the special rates of postage prescribed by Section 132.122.

Printed in United States of America

Number of copies of this issue 5,500

Limit-Cycle Stability Study of a Feedback Control System by a New Describing-Function Technique

H. J. HARRINGTON
ASSOCIATE MEMBER AIEE

A NEW describing-function technique is used in this paper to show the ects of several gain-controlling non- earities on the limit-cycle stability of an craft roll-stabilization system. The w method leads to a stability diagram ich is divided not into the usual two gions of "stable" and "unstable" opera- n but rather into numerous regions each which is associated with a particular pe of limit-cycle oscillation or singular int. Such a diagram gives considerable ight into the mechanism of oscillations en observed in physical equipment. ry satisfactory agreement with analog mputer studies is demonstrated for the w technique.

The case of simultaneous limiting in h of two feedback loops is easily handled the system studied herein. A logical thesis procedure for deliberately in- ducing a new nonlinearity in order to nteract an existing one is also pro- sed.

A paper[1] treating the stability of linear tems furnished the impetus for this rk; another paper[2] concerned itself h some of the same problems which treated here but the approach and erpretation of results were considerably ferent.

Although the technique developed here y be employed for any nonlinearity for ich a describing function can be derived, appears to offer interpretative ad- ntages over the more conventional ap- ach[3] only in the case of gain-controlling nlinearities; that is, nonlinearities ich do not introduce phase shift to nals passing through them.

er 59-233, recommended by the AIEE Feedback trol Systems Committee and approved by the B Technical Operations Department for entation at the AIEE Winter General Meeting, York, N. Y., February 1–6, 1959. Manuscript mitted September 29, 1958; made available for ting December 22, 1958.

J. HARRINGTON was formerly with Convair- onautics, A Division of General Dynamics poration, San Diego, Calif., and is now with r Aircraft Company, San Diego, Calif.

author wishes to thank Mr. Carl J. Scherer McDonnell Aircraft Corporation for first using methods of this paper in a practical application also for obtaining the analog computer checks ented herein.

Nomenclature

A = peak amplitude of a sinusoidal oscillation in roll rate
d = value of threshold in rate gyroscope
$E_m = M_d - M_i$, error in normalized rolling moment
f = viscous damping coefficient of air frame
I = moment of inertia of air frame about roll axis
i = control surface incidence
i_c = control surface incidence command
$j = \sqrt{-1}$
$K_a = M_i/i$, aerodynamic effectiveness of control surface
K_p = position feedback gain
K_r = rate feedback gain
L_p = position-gyro limit
L_r = rate-gyro limit
M_d = rolling moment (normalized to I) caused by external disturbance
M_i = rolling moment (normalized to I) caused by control surface incidence
N_p = equivalent gain, or describing function, of nonlinearity in position gyro
N_r = equivalent gain, or describing function, of nonlinearity in rate gyro
s = Laplace operator
T_1, T_2 = constants of control servo
$\alpha_d = d/A$, amplitude parameter for threshold describing function
$\alpha_l = L_r/A$, amplitude parameter for limiting describing function
$\gamma = L_p/L_r$, position-to-rate limit ratio
ϕ = roll-position angle
$\Omega = \gamma\omega$, normalized frequency
ω = circular frequency

Description of System Studied

The system to be studied, which is shown in block diagram in Fig. 1, is a typical aircraft roll-stabilization system used to maintain air-frame roll angles near zero in the presence of external disturbances. The system works as follows. A disturbing moment M_d produces a moment error which in turn causes the air-frame to roll. The instantaneous roll rate and position are measured by gyroscopes whose output signals are suitably scaled by the gains K_r and K_p, added, and sent to a control surface servo as an incidence command i_c. The output incidence of the control servo then gives rise to a control moment M_i, which tends to reduce the moment error and thus stabilize the air-frame.

This seemingly simple system is complicated by the presence of limiting in, or immediately following, the position gyro and by both limiting and threshold in the rate gyroscope. In addition, the aerodynamic gain K_a is subject to rather wide variations from nominal because of unpredictable changes in aircraft flight conditions; good stability margins with respect to the gain K_a must therefore be attained to prevent either divergent oscillations or a hard limit cycle of catastrophic proportions.

In the study which follows, it will be assumed that the linearized transfer function of both gyroscopes is unity. Although this might well be a good assumption, the restriction is not at all necessary and is made only for the sake of brevity as is the omission of compensating networks in either of the two feedback paths. The nominal system parameters for the rest of the system are as follows:

$$\begin{matrix} f/I = 0.67 \text{ per second} \\ T_1 = 0.04 \text{ second} \\ T_2 = 0.05 \text{ second} \end{matrix} \quad \left.\begin{matrix} K_a & d \\ K_p & L_p \\ K_r & L_r \end{matrix}\right\} \text{general} \qquad (1)$$

Generalized Stability Chart

The characteristic equation $F(s)$ of the system of Fig. 1 is

$$F(s) = T_1^2 s^4 + \left(T_2 + T_1^2 \frac{f}{I}\right)s^3 + \left(1 + T_2 \frac{f}{I}\right)s^2 + \left(K_a K_r N_r + \frac{f}{I}\right)s + K_a K_p N_p \qquad (2)$$

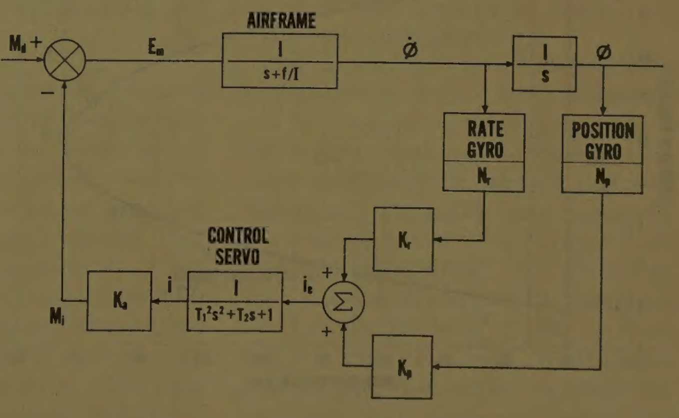


Fig. 1. Block diagram of an aircraft roll-control system

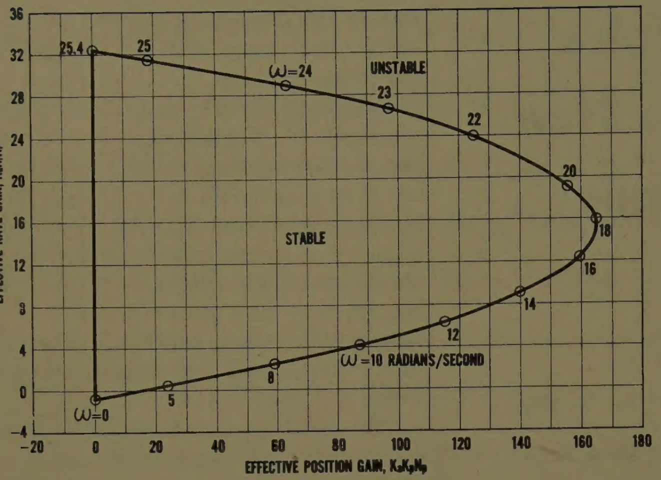


Fig. 2. Stability diagram for the roll-control system

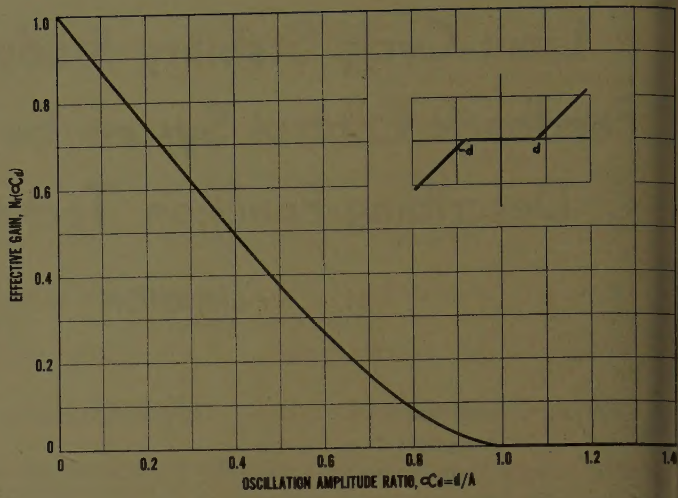


Fig. 3. Describing function for threshold

For the present the nonlinearity terms N_p and N_r are assumed to be simple constants. Setting equation 2 equal to zero and replacing s by $j\omega$ gives an equation which represents the marginally stable case if real roots in ω exist:

$$T_1{}^2\omega^4-\left(1+T_2\frac{f}{I}\right)\omega^2+K_aK_pN_p-j\omega\left[\left(T_2+T_1{}^2\frac{f}{I}\right)\omega^2-\left(K_aK_rN_r+\frac{f}{I}\right)\right]=0 \quad (3)$$

But for equation 3 to be valid, both the real and imaginary parts must be identically zero; therefore, this equation may be replaced by two separate equations from which the system gains can be solved

$$K_aK_pN_p=\left(-T_1{}^2\omega^2+1+T_2\frac{f}{I}\right)\omega^2 \quad \text{(real part)} \quad (4)$$

$$K_aK_rN_r=\left(T_2+T_1{}^2\frac{f}{I}\right)\omega^2-\frac{f}{I} \quad \text{(imaginary part)} \quad (5)$$

For the particular parameter values in equation 1, these equations become

$$K_aK_pN_p=(-0.0016\omega^2+1.033)\omega^2 \quad (6)$$

$$K_aK_rN_r=0.0511\omega^2-0.67 \quad (7)$$

A plot of equations 6 and 7 with ω treated as a real-value parameter is shown in Fig. 2. By applying Hermite's criteria[1] (or any other convenient method), it can be shown that the region within the closed boundary is stable while that outside the boundary line is unstable. This situation is approximately true even when the system is nonlinear if the effective gains of the nonlinearities N_p and N_r are explicitly included in the gain co-ordinates as shown in Fig. 2.

Limit-Cycle Stability Chart

For better interpretation of the nonlinear cases, it is desirable to have the stability plot in terms of just the linear gains K_aK_p and K_aK_r, and show not only the wholly stable and unstable regions but also the various limit-cycle oscillation conditions which might arise. To illustrate this, the case of a single nonlinearity, namely, threshold in the rate gyro, will be examined.

It will be assumed that the effective gain of the threshold element N_r is given by its describing function[3] during an oscillation. A plot of N_r as a function of normalized roll-rate amplitude α_d is shown in Fig. 3. Note that if the roll-rate amplitude is equal to or less than the deadspot, that is $\alpha_d \geq 1$, the gain N_r is zero. As the roll-rate amplitude approaches infinity, however, the grain N_r approaches unity and the system should behave linearly.

The consequences of this nonlinearity may be seen by examining Fig. 4. Here the solid curve is identical to that of Fig. 2 while the broken lines, which reflect the presence of the threshold, divide the over-all chart into regions of varying degrees of stability. In general, linear gain co-ordinates appearing as a point in Fig. 4 will be transformed into a vertical line in Fig. 2. Since N_p is by definition unity, this line will be located at the same abscissa and will descend from the point ordinate ($N_r=1$) to the position-gain axis.

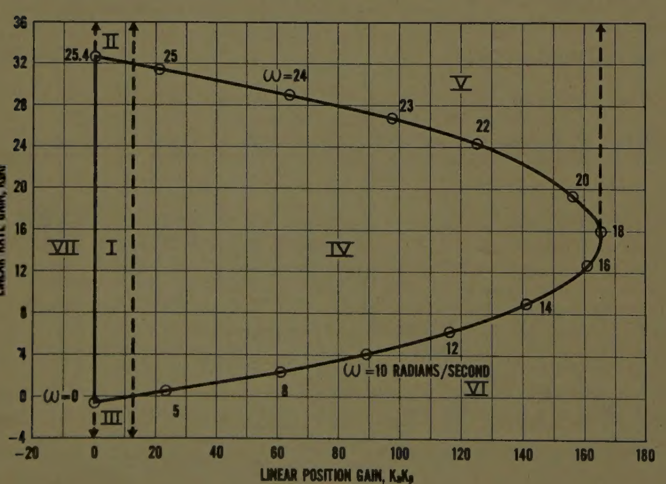
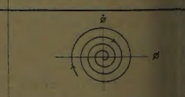
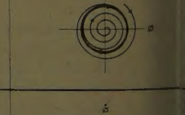
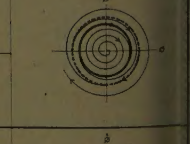
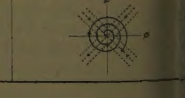


Fig. 4 (left). Limit-cycle stability diagram for case of threshold in rate gyroscope

Fig. 5 (right). Phase-plane sketches showing typical trajectories associated with the seven regions of Fig. 4

REGION	FREQ. RAD./SEC.	STABILITY DESCRIPTION	PHASE-PLANE SKETCH
I	—	STABLE SINGULAR POINT	
II	>25.2 <25.4	HARD, UNSTABLE LIMIT CYCLE	
III	>0 <3.6	HARD, UNSTABLE LIMIT CYCLE	
IV	>3.6 <18	SOFT, STABLE LIMIT CYCLE	
V	>3.6 <18	SOFT, STABLE LIMIT CYCLE	
V	>18 <25.2	HARD, UNSTABLE LIMIT CYCLE	
VI & VII	—	UNSTABLE SINGULAR POINT	

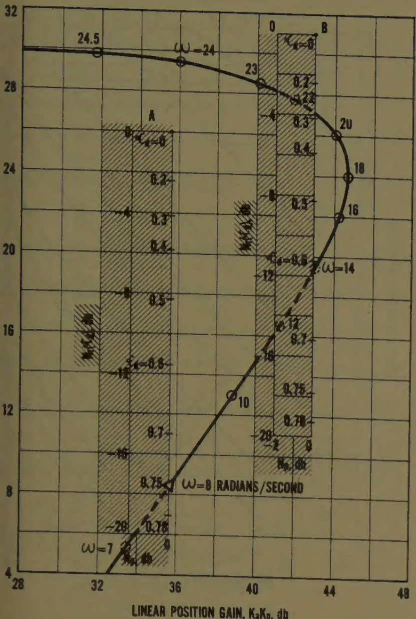


Fig. 6 (left). Logarithmic plot of limit-cycle stability diagram showing overlay of describing function for rate-gyro threshold

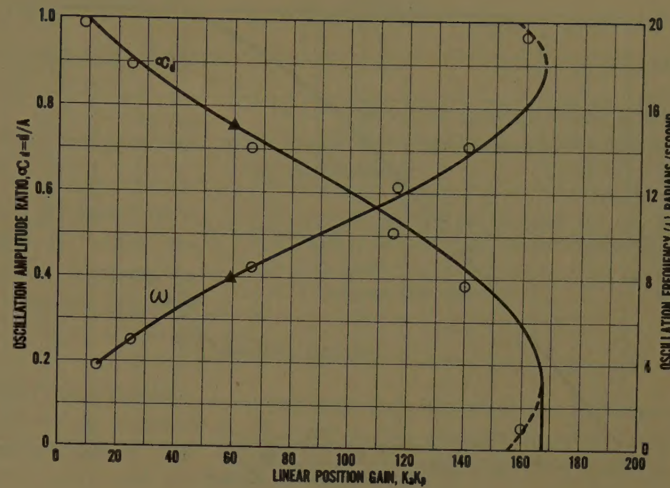


Fig. 7. Amplitude and frequency of limit-cycle oscillations caused by rate-gyro threshold for variations in position gain K_aK_p, while holding rate gain constant at $K_aK_r=20$

—— Stable oscillations as determined by describing functions
– – – Unstable oscillations as determined by describing functions
O Points obtained by analog computer

$N_r=0$). If the linear-gain co-ordinates ıll in region I of Fig. 4, no amount of ttenuation because of $N_r(\alpha_d)$ can cause ıe corresponding gain ordinate in Fig. 2 fall in the unstable region. Thus, for gion I, any initial disturbance must ltimately die out as is depicted by the orresponding phase-plane sketch of Fig. (The phase-plane sketches of Fig. 5 e for purposes of qualitative description ıly since a polydimensional phase space required to accurately describe higher-der systems.)

If the linear-gain co-ordinates fall in ther regions II or III, it is possible to cite an unstable limit-cycle oscillation. his behavior may be predicted by noting om Fig. 2 that the system is stable when rest ($N_r=0$), and unstable when excited fficiently to cause the gain $K_aK_rN_r$ to ceed the stable-region ordinate.

Region IV of Fig. 4, which is the most obable region of operation for the stem in question, leads to a soft, stable nit cycle. For examination of Figs. 2 d 3 shows that with the system at rest e effective rate gain is zero and the system is unstable; but, as the amplitude the resulting divergent oscillation ilds up, α_d decreases and $N_r(\alpha_d)$ increases until the gain reaches the lower ability boundary of Fig. 2. Further creases in the oscillation amplitude force e effective gain ordinate into the stable gion with the result that the excess nplitude is damped out. Obviously the quency of oscillation must lie in the nge of $3.6<\omega<18$ radians per second.

Both a soft, stable limit cycle and a gher-amplitude unstable oscillation are ssible if the system gains are adjusted lie in region V of Fig. 4. Repetition the preceding logic will show that the stable oscillation will occur at a frequency corresponding to the lower stability boundary of Fig. 2 while the unstable limit cycle will be at one of the higher frequencies of the upper stability boundary.

Finally, regions VI and VII represent totally unstable conditions for the case of rate-gyro threshold just as they do for a linear system since for these gains no realizable value of $N_r(\alpha_d)$ can lead to the stable region of Fig. 2. As shown in Fig. 5, the phase-plane sketch characterizing these regions consists of typical trajectories near an unstable singular point.

The preceding discussion has shown that the stability plot for the linear system (Fig. 2, $N_p=N_r=1$) consists of only two separated regions which define the system as being either stable or unstable. The corresponding limit-cycle stability chart for the case of rate-gyro threshold, however, is divided into seven separated regions which represent five distinct types or degrees of stability.

Quantitative Limit-Cycle Data for Threshold

So far, only the nature of possible limit-cycle oscillations caused by threshold have been discussed. Now the amplitude and frequency of the oscillations will be quantitatively determined and compared with the results of analog computer tests.

The normalized amplitude α_d of the limit cycle can always be determined by computing the attenuation necessary for oscillations from Fig. 4 and then referring back to Fig. 3 to see what α_d value gives this attenuation. A more convenient and elucidating method, however, is to replot the portions of the stability chart and describing function of interest to a logarithmic scale; the describing-function graph can then be used as an overlay and the gain division indicated previously becomes simple graphical subtraction.

As an illustration, portions of Figs. 3 and 4 are replotted in terms of decibels (db) in Fig. 6. Here, the describing-function graph is drawn as $N_r(\alpha_d)$ versus N_p with α_d as a parameter; since the position gyro is presently assumed linear, N_p is identically unity (zero db) for all values of α_d. Two separate overlays of the describing function are shown on the stability chart of Fig. 6 as they would appear for the following gain conditions

Case A: $K_aK_r=26$ db; $K_aK_p=35.6$ db
Case B: $K_aK_r=31$ db; $K_aK_p=43$ db

For each of these system gains, the oscillation amplitude and frequency may be immediately determined by noting the values of α_d and ω corresponding to the intersections of the two curves. Note that for case B, both a stable and unstable solution exists for respectively the lower and upper intersection.

The accuracy with which the amplitude and frequency of limit-cycle oscillations can be predicted is illustrated in Fig. 7 which compares the results of analog computer studies and the proposed analytical technique. The analog setup employed a conventional biased-diode circuit for the threshold simulation; all harmonics from this circuit were allowed to circulate freely around the loop. The system rate gain was held at $K_aK_r=20$ (26 db) and the position gain then varied to obtain the amplitude and frequency points and

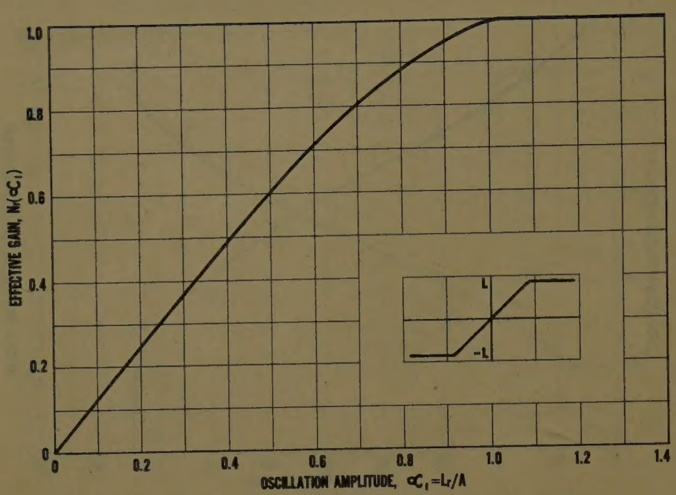
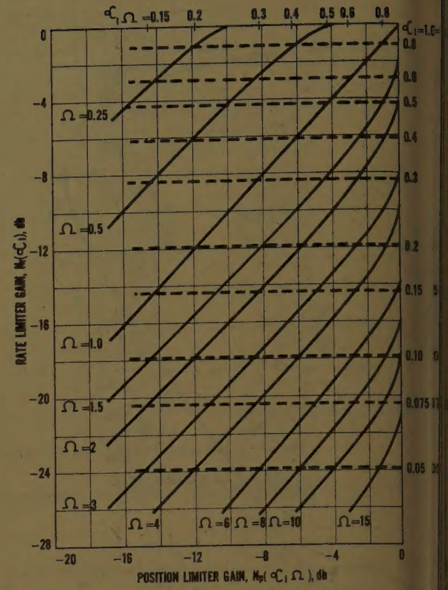


Fig. 8 (left). Describing function for limiting

Fig. 9 (right). Describing - function overlay for simultaneous limiting in position and rate gyros. Each curve is associated with a single value of normalized frequency $\Omega = \gamma\omega$



curves shown in Fig. 7. The darkened triangles appearing on the analytically determined curves correspond to the similarly marked point in Fig. 6, case A.

It may be noted from Figs. 4 and 5 that the gain conditions $K_aK_r=20$, $156<K_aK_p<167$ corresponds to region V wherein both stable and unstable oscillations are predicted. Both types were verified by the analog computer study and, as shown in Fig. 7, the proposed analytical method is capable of determining the amplitude and frequency of either type oscillation with quite suitable engineering accuracy. (It would appear, however, that the same reservations concerning higher harmonics applicable to conventional describing-function theory[3] are equally applicable to the new method.)

Rate-Gyro Limiting

The effects of limiting in the rate gyro can be determined by employing the same procedure just discussed for threshold after substituting the now-appropriate describing function for limiting shown in Fig. 8. Such an investigation reveals that the limit-cycle stability diagram remains divided into the same seven regions shown in Fig. 4. In all but region V, however, the stable and unstable oscillations shown in Fig. 5 become interchanged in going from threshold to limiting; in region V the frequencies of the two limit cycles become interchanged.

Analog-computer studies of the limit cycles resulting from rate-gyro limiting were also conducted. The analytical results were verified by this means to within essentially the same accuracy as that shown in Fig. 7 for threshold.

Investigation of the rate-gyro limiting case leads to a very interesting conclusion. The whole of region IV, Fig. 4, is plagued by a hard limit-cycle oscillation and yet this is the only region of practical operation from a linear standpoint. Since the threshold of a physical rate gyro usually increases in direct proportion to the maximum range, or limit value, it is usually not wise to specify a very large limiting value in an attempt to rationalize the limit cycle out of existence. What, then, can be done to circumvent this problem area? One answer is to deliberately introduce another nonlinearity into the system.

Rate- and Position-Gyro Limiting

It is obvious that position-gyro limiting would cause the describing-function loci of Fig. 6 to swing towards the left, for the value of N_p would then become less than unity. This suggests that operation in region IV, Fig. 4, might be possible without the potential hard limit cycle if position-gyro limiting were employed.

To investigate this possibility, assume that an oscillation in roll exists such that

$$\dot{\phi}=A \sin \omega t \tag{8}$$

$$\phi=\int \dot{\phi} dt=-\frac{A}{\omega} \cos \omega t \tag{9}$$

If now the rate signal is limited to the value $\pm L_r$, and the position signal is limited to $\pm L_p$, the describing function $N(\alpha)$ for each limiter will be as shown in Fig. 8; here, α is defined for the rate limiter as

$$\alpha=\frac{L_r}{A}=\alpha_1 \tag{10}$$

and for the position limiter,

$$\alpha=\frac{L_p}{A/\omega}=\frac{L_r}{A}\frac{L_p}{L_r}\omega=\alpha_1\gamma\omega=\alpha_1\Omega \tag{11}$$

From equations 10 and 11, a new overlay chart of $N_r(\alpha_1)$ versus $N_p(\alpha_1\Omega)$ may be constructed with loci for several values of normalized frequency, $\Omega=\gamma\omega$, as shown in Fig. 9. This overlay can be used in conjunction with a linearized-gain stability plot just as was done in Fig. 6 except now only those intersections occurring at like frequencies are valid. If a region can be found in which no valid intersections occur, then no limit cycles due to gyro limiting are possible in this region.

The determination of such a region is demonstrated in Fig. 10 which shows a few values of the $N_r(\alpha_1)$ versus $N_p(\alpha_1\Omega)$ overlay drawn on the stability chart for the special case of $\gamma=L_p/L_r=1/5$ second.

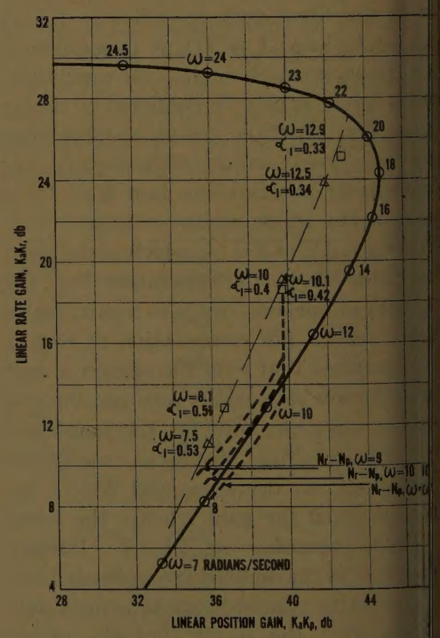
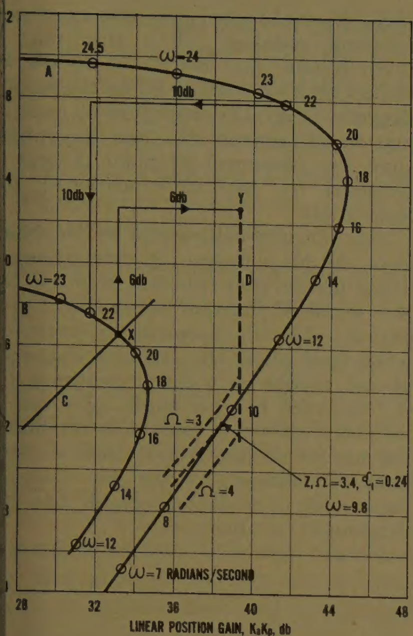


Fig. 10. Logarithmic plot of stability diagram and describing-function overlay for simultaneous gyro limiting in the ratio of $\gamma=1$

– – – Separates regions of stable and hard limit cycle operation

△ Boundary solutions obtained by describing function

□ Solutions obtained from analog computer



11. **Logarithmic plot of stability diagram showing synthesis procedure**

te that if the describing-function over- were moved up from the gain value wn, no more valid (like-frequency) ersections would occur. But if the n were lowered, two proper intersec- ns would result: one corresponding to ard, unstable limit cycle of frequency 10.6 radians per second, and the other table oscillation of somewhat greater plitude and lower frequency. It is y apparent that as the system gain proaches a value for which valid inter- ions are only marginally possible, the limit cycles coalesce to destroy one ther.[4] Such a gain, therefore, repre- ts one point on a boundary line which arates the stability chart into regions table and potentially unstable (hard t cycle) operation.

he heavy dashed line of Fig. 10 shows analytically determined locus of sys- gains resulting in neutral limit-cycle bility for $\gamma=1/5$ second. Also shown comparison are the amplitude and uency of several of these marginal llations as determined both ana- cally and by the analog computer. e the excellent agreement in all ects.

ynthesis Procedure

aving developed the tools necessary andle the nonlinearities in the system, pose now that it is desired to determine best gains and gyro limit values to sfy the following design requirements:

The feedback gain ratio shall be ad- ed such that $K_r/K_p=0.15$ second.

The linear-system gain margin with ect to K_a shall be 10 db.

3. The hard limit-cycle gain margin with respect to K_a shall be 6 db.

4. The system shall be capable of measuring roll angles up to ±25 degrees.

5. The amplitude of soft, limit-cycle oscillations due to rate-gyro threshold shall be held to a minimum.

The first requirement is based on a linear root locus study in which the amount of rate-to-position feedback required to provide an adequate transient response in the linear mode of operation was determined.

In the second and third requirements, the gain margin is expressed in terms of K_a since, as mentioned previously, this gain is subject to rather wide and unpredictable variations. The limit-cycle gain margin is specified somewhat lower than that for the linear case since not only the system gain but also the level of excitation existing simultaneously must be favorable before oscillations can develop. The relative weighting of the two gain margins is largely a matter of judgement.

The fourth and fifth requirements state, in essence, that the position-gyro limit be set at 25 degrees. For if this limit were higher than necessary, the rate-gyro limiter would also have to be higher to provide a given limit-ratio value γ, and this in turn would increase the threshold level of the rate gyro because of physical construction considerations.

The synthesis procedure is shown in Fig. 11. First, the linear frequency locus is plotted as curve A. Next, curve B is established by projecting each point of curve A back and down 10 db; obviously curve B represents a boundary outside of which the linearized gain margin with respect to K_a is less than 10 db. The condition for $K_r/K_p=0.15$ is depicted as the straight line C. Thus the point X satisfies the first two requirements previously listed and represents the desired gain condition.

By virtue of requirement 3, the point Y may be determined as falling on the limit-cycle stability boundary; therefore the describing-function overlay, curve D, must furnish a marginally stable solution when its zero-zero db point is aligned with Y. This marginally stable solution is shown as point Z in Fig. 11 and occurs for $\omega=9.8$ radians per second and $\Omega=\gamma\omega=3.4$. From these data, the limit ratio γ may be solved for as

$$\gamma=\frac{L_p}{L_r}=\frac{\Omega}{\omega}=\frac{3.4}{9.8} \text{ second} \tag{12}$$

Now from requirement 4 and equation 12, the desired value of rate limit, L_r, becomes

$$L_r=\frac{L_p}{\gamma}=\frac{(25)(9.8)}{3.4}=72 \text{ degrees per second} \tag{13}$$

In summary, the system which meets the specified requirements has the following parameters:

$K_aK_r=16.6$ db $=6.8$ per second
$K_aK_p=33.2$ db $=46$ per second squared
$L_p=25$ degrees
$L_r=72$ degrees per second

Conclusions

The new describing-function technique presented herein appears to offer quite substantial advantages in interpretation and flexibility over the more conventional approach to limit-cycle analysis when the system nonlinearities are of the gain-controlling type. A slight modification[2] would allow phase-shifting nonlinearities to be handled also but with dubious benefits.

The additional utility of the new method stems from having two generalized system parameters with which to work rather than the single gain associated with open-loop diagrams.[3] The price paid for this convenience is the somewhat laborious derivation and computation of the characteristic equation of the closed-loop system, particularly if the order of the equation is high. To obtain the same amount of information and insight by the conventional method would, however, lead to vastly more labor since families of open-loop frequency loci would be necessary to account for the second parameter.

For the system analyzed in this paper, it was convenient to choose two gains for the co-ordinates of the stability diagram. In other problems it might be more appropriate to choose one gain and some other system parameter,[1] a time constant for example. In other words, the method is considerably more flexible than is demonstrated by the particular analysis carried out herein.

References

1. A New Stability Criterion of Linear Servomechanisms by a Graphical Method, Toshio Numakura, Takeo Miura. *AIEE Transactions*, vol. 76, pt. II (*Applications and Industry*), Mar. 1957, pp. 40–48.

2. Analyses of Servomechanisms by Describing Function Method on a Plane with Parameters as Coordinates, Toshio Numakura, Takeo Miura, Akira Kamoi. *AIEE CP57-1035* (available on request).

3. Sinusoidal Analysis of Feedback-Control Systems Containing Nonlinear Elements, E. Calvin Johnson. *AIEE Transactions*, vol. 71, pt. II (*Applications and Industry*), July 1952, pp. 169–81.

4. Non-Linear Mechanics (book), N. Minorsky. J. W. Edwards, Inc., Ann Arbor, Mich., 1947, pp. 68–73.

Discussion

Toshio Numakura (Hitachi Central Research Laboratory, Kokubunji, Tokyo, Japan): The author has solved a complicated system which contains several nonlinearities by using the stability diagrams introduced in references 1 and 2 of the paper.

The new ideas added to the describing-function method and the works cited are very interesting.

Some problems may remain unsolved in the last section of the paper, "A Synthesis Procedure," in which the author intends to predict the transient response of the nonlinear system by use of the stability diagram. Fig. 11 does not provide any indication of the phase margin or slope of the gain characteristic as does Bode's diagram, so it is not possible to predict the transient response exactly. Though the discusser and his colleagues intended to find a suitable method by which the transient response could be predicted, they failed to find any not only in the case of the nonlinear system but for the linear system as well. Therefore the optimum parameter conditions should be found by using the stability diagram in conjunction with an analog computer.

H. J. Harrington: The author appreciates the comments of Dr. Numakura and the resulting opportunity to elaborate on the paper.

Actually, there is no intent in the paper to predict the transient response by means of the stability diagram. Thus, in the section referred to by the discusser, it is pointed out that a complementary root-locus analysis furnished the gain ratio value of $K_r/K_p=0.15$ required to give an adequ transient response. It is true that t linear-system gain margin of 10 db w respect to K_a is marked off on the stabili diagram (Fig. 11), but this is mostly matter of convenience for presentati since the placement of point X was c termined by the afore-mentioned ro locus study.

The difficulty experienced by Dr. Num kura and his colleagues in determining t phase margin and transient response fr the stability diagram is shared by t author. This limitation is so severe th when used alone the stability diagram almost useless as a linear synthesis to As expanded by this paper, however, t method is extremely valuable in establishi the nature of steady-state limit cycles mildly nonlinear systems in which t linear-range transient response has be determined by other means.

New Methods of Simplifying Boolean Functions

R. L. HOWARD
ASSOCIATE MEMBER AIEE

ECONOMY IS ONE of the most important considerations in the design of any system. Maximum savings can be achieved in the design of switching circuits by performing the desired logic in the simplest minimum form applicable to the equipment utilized. Boolean algebra is often used to express the functions of switching circuits and these Boolean expressions generally appear in the final form of a sum of products or a product of sums. The form that presents the most economical application is determined by the equipment. It is therefore necessary to be able to transform and simplify Boolean functions so that they will appear in the predetermined form that will yield the maximum savings.

Methods for performing the simplification and transformation of Boolean functions have been presented in previous publications by Karnaugh,[1] Quine,[2] and McCluskey.[3] Each of these processes becomes extremely complicated for complex functions. Furthermore, some duplications are encountered which repeat information during the process that initially had been stated in the original function.

The following discussion describes a new procedure of simplifying and transforming any Boolean function without some of the restrictions encountered in previously presented procedures. The original Boolean function may consist of any number of variables and may be expressed in any form (sum of product, product of sums or a combination of both). Without changing the form of the original function the same basic rules apply when either simplification and transformation is made; only the recording of the original function is changed. This provides a systematic process for the reduction of static or conventional control systems and also an easily programmed process for digital computer application.

Discussion

For clarity the explanation of the method includes a sample Boolean function. Wherever the sample is concerned in the explanation that portion will be labelled "sample." Also included wherever possible, are examples which are independent of the sample function. Each of these will be labelled "Example."

The sample Boolean function is

$$X=A\cdot B[C\cdot D\cdot E+A\cdot \bar{C}\cdot \bar{E}(C+D)]$$
$$=AB[CDE+A\bar{C}\bar{E}(C+D)]$$

where

$+$ = OR function
$\cdot$ = AND function

The "variables" of a function are the number of different letters concerned. The "literals" of a function are the individual entries in the function. Th the sample function contains five va ables and ten literals. Fig. 1 sho logic diagrams of the sample function.

The Basic Procedure of the Meth

Preparation of the Function for Transformation and Simplification

1. Determine the number of variab contained in the original expression.

Sample: There are five variables contain in the expression (A, B, C, D and E).

2. Assign a digit to each variable.

	Digits				
Sample:	E	D	C	B	A

3. The digital representation of ea product is now written as follows:

If the literal indicates the presence of variable, a "1" will be placed in the c responding digit for that term.

If the literal indicates the absence (negati of the variable, a "0" will be placed in corresponding digit for that term.

If the literal does not appear in the term dash (–) will be placed in the correspondi digit for that term.

Sample: $AB[CDE+A\bar{C}\bar{E}(C+D)]$

E	D	C	B	A		E	D	C	B	A
–	–	–	1	1		[1	1	1	–	–

E	D	C	B	A		E	D	C	B	A
0	–	0	–	1		(–	–	1	–	–

E	D	C	B
–	1	–	–

4. For the conversion of the origi

Paper 59-91, recommended by the AIEE Indust Control Committee and approved by the A Technical Operations Department for presenta at the AIEE Winter General Meeting, New Y N. Y., February 1–6, 1959. Manuscript mitted November 3, 1958; made available printing December 16, 1958.

R. L. Howard is with Westinghouse Ele Corporation, Pittsburgh, Pa.

ession into the sum of products ich will be referred to as the AND): (*a*). Wherever an AND function is cated in the original expression the two ns it concerns will be combined. (*b*). erever an OR function is indicated in original expression the two terms it erns will not be combined.

For the conversion of the original ex- sion into the product of sums (which be referred to as the OR form): (*a*). erever an OR function is indicated in original expression the two terms it erns will be combined. (*b*). Wherever ND function is indicated in the original ession the two terms it concerns not be combined.

Symbols used for indicated combina- and noncombinations are a long (—) between two terms to indicate they are to be combined; and a long ed dash (-+-) used between two s to indicate that they are not to be ined. These symbols will vary in h to correspond to the original ession.

The original expression is recorded itten). The digit representation of term is written, one term beneath other. The proper indications for ination or noncombination are in- d in the following vertical representa- of the expression. Note that the l representation of the expression ins the same for both transforma- . The only difference is that the ination and noncombination indica- are reversed.

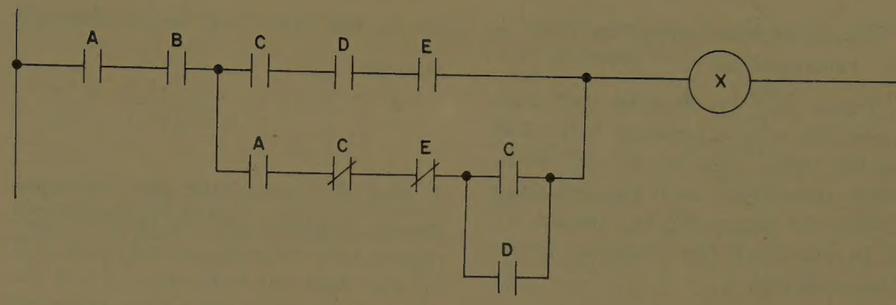
le:

C	B	A		E	D	C	B	A
–	1	1		[1	1	1	–	–+

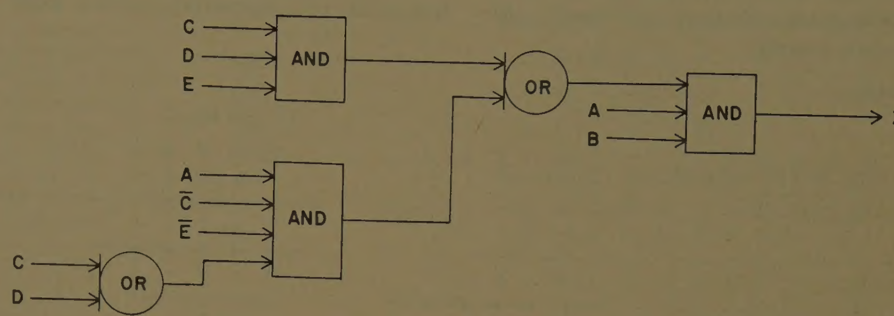
E	D	C	B	A		E	D	C	B	A
0	–	0	–	1		(–	–	1	–	–+
						E	D	C	B	A
						–	1	–	–	–)]

To AND Form

```
E D C B A
– – – 1 –
─────────
– – – – 1
─────────────────
1 – – – –
─────────
– 1 – – –
─────────
– – 1 – –
─+─+─+─+─+─
0 – – – –
─────────
– – 0 – –
─────────
– – – – 1
─────────────
– – 1 – –
+─+─+─+─+
– 1 – – –
```



(A)



(B)

Fig. 1. Sample expressions

A—Relay diagram
B—Static diagram

To OR Form

```
E D C B A
– – – 1 –
+─+─+─+─+
– – – – 1
───+─+─+─+─+───
1 – – – –
+─+─+─+─+
– 1 – – –
+─+─+─+─+
– – 1 – –
─────────────
0 – – – –
+─+─+─+─+
– – 0 – –
+─+─+─+─+
– – – – 1
──+─+─+─+─+──
– – 1 – –
─────────
– 1 – – –
```

Rules for Transforming the Function

1. Rules for any digit of two terms that are to be combined are:

(*a*). A 0 in the digit of one term and a 1 in the corresponding digit of the other term. The two terms cannot be combined and there will be no resulting term.

Example: Digit

$$\frac{1}{0} = \text{no resulting term}$$

(*b*). A 0 in the digit of one term and a 0 in the corresponding digit of the other term will yield a 0 in the same digit of the resulting term.

Example: Digit

$$\frac{0}{0} = 0$$

(*c*). A 0 in the digit of one term and a dash (–) in the corresponding digit of the other term will yield a 0 in the same digit of the resulting term.

Example: Digit

$$\frac{0}{-} = 0$$

(*d*). A 1 in a digit of one term and a 1 in the corresponding digit of the other term will yield a 1 in the same digit of the resulting term.

Example: Digit

$$\frac{1}{1} = 1$$

(*e*). A 1 in a digit of one term and a dash (–) in the corresponding digit of the other term will yield a 1 in the same digit of the resulting term.

Example: Digit

$$\frac{1}{-} = 1$$

(*f*). A dash (–) in a digit of one term and a dash (–) in the corresponding digit of the other term will yield a dash (–) in the same digit of the resulting term.

Example: Digit

$$\frac{-}{-} = -$$

2. Noncombinable terms are left in their existing form.

3. Terms may be grouped that have consecutive noncombination indications between them. The terms will be repeated (rewritten) and the indication between the terms will be omitted. It will be understood that the terms are not to be combined.

4. The indicated combination concerning the smallest number of terms will be performed first. The successively larger combinations are then performed until no more remain.

one or the other may be eliminated.

Example:

```
E D C B A     E D C B A
- 1 1 - 1  =  - 1 1 - 1
- 1 1 - 1
```

2. For two terms that have the same number of literals and the identical digit representation except for one digit. If for that digit one term has a 0 and the other a 1, the two terms may be expressed as one new term. This new term will duplicate the common digit representation of the two orginal terms and a dash

tains a lesser number of literals than th other and also has its literal digit repre sentation duplicated in the other term The term with the greater number c literals is redundant and may be elim inated.

Example:

```
E D C B A     E D C B A
- 1 0 1 -  =  - 1 0 1 -
- 1 0 1 1
```

4. For two terms, one of which contai a lesser number of literals than th other and also has its literal digit repr

Sample:

To AND Form

```
E D C B A      E D C B A      E D C B A      E D C B A
- - - 1 -      - - - 1 1      - - - 1 1      - - - 1 1
---------      ---------      ---------      ---------
- - - - 1      1 1 - - -      1 1 1 - -      1 1 1 - -
---------      ---------      -+-+-+-+-      -+-+-+-+-
1 - - - -      - - 1 - -      0 - 0 - 1      0 1 0 - 1
---------      -+-+-+-+-      ---------
- 1 - - -      0 - 0 - -      - - 1 - -      E D C B A
                                             1 1 1 1 1
---------      ---------
- - 1 - -      - - - - 1      - 1 - - -      0 1 0 1 1
-+-+-+-+-      ---------
0 - - - -      - - 1 - -
---------
- - 0 - -      - 1 - - -
---------
- - - - 1
---------
- - 1 - -
-+-+-+-+-
- 1 - - -
```

To OR Form

```
E D C B A      E D C B A      E D C B A      E D C B A      E D C B A
- - - 1 -      - - - 1 -      - - - 1 -      - - - 1 -      - - - 1 -
-+-+-+-+-
- - - - 1      - - - - 1      - - - - 1      - - - - 1      - - - - 1
-+-+-+-+-      -+-+-+-+-      -+-+-+-+-      -+-+-+-+-
1 - - - -      1 - - - -      1 - - - -      0 1 - - -      0 1 - - -
-+-+-+-+-
- 1 - - -      - 1 - - -      - 1 - - -      0 - 1 - -      0 - 1 - -
-+-+-+-+-
- - 1 - -      - - 1 - -      - - 1 - -      1 - 0 - -      1 - 0 - -
---------      ---------      ---------
0 - - - -      0 - - - -      0 - - - -      - 1 0 - -      - 1 0 - -
-+-+-+-+-
- - 0 - -      - - 0 - -      - - 0 - -      1 - - - 1      1 - - - 1
-+-+-+-+-
- - - - 1      - - - - 1      - - - - 1      - 1 - - 1      - 1 - - 1
-+-+-+-+-      -+-+-+-+-
- - 1 - -      - 1 1 - -      - 1 1 - -      - - 1 - 1      - - 1 - 1
---------
- 1 - - -                                    1 1 1 - -      1 1 1 - -
                                             - 1 1 - -      - 1 1 - -
                                             - 1 1 - -      - 1 1 - -
```

Rules for Simplifying the Functions

A literal will be considered to be contained in a digit if that digit has a 0 or 1 in it. The literal digit representation of a term is the 0 and/or 1 configuration contained in its digits. If a dash (–) is present in the digit the literal is not contained in that digit.

1. If two identical terms are repeated,

(–) will be placed in the digit that varied from 0 to 1 (or 1 to 0).

Example:

```
E D C B A     E D C B A
- 1 - 1 1  =  - 1 - - 1
- 1 - 0 1
```

3. For two terms, one of which contains a lesser number of literals than the

sentation duplicated except for one d in the other term, if for that one digit term has a 0 and the other a 1, the te containing the least number of lite is retained. The other term become new term that duplicates the comm digit representation of the two or nal terms and a dash (–) is placed

digit that varied from a 0 to 1 (or 0).

ample:

D	C	B	A		E	D	C	B	A
1	1	–	1	=	–	1	1	–	1
1	0	1	1		–	1	–	1	1

The sample for the AND form used in m *4* under "Rules for Transforming the nction," indicates that the remaining o terms cannot be simplified and there- e constitute the final expression.

Sample:

E	D	C	B	A	
–	–	–	1	1	P
1	1	1	–	–	P
0	–	0	–	1	P
–	1	1	–	–	S

(*c*). The logic functions between terms will be handled exactly as was outlined in items *4* through *7* under "The Basic Procedure of the Method."

	E	D	C	B	A
)	–	–	–	1	–
)	–	–	–	–	1
)	0	1	–	–	–
)	0	–	1	–	–
)	1	–	0	–	–
)	–	1	0	–	–
)	1	–	–	–	1
)	–	1	–	–	1
)	–	–	1	–	1
)	1	1	1	–	–
)	–	1	1	–	–
)	–	1	1	–	–

To OR Form

	E	D	C	B	A
(2)	–	–	–	1	–
(4)	0	1	–	–	–
(5)	0	–	1	–	–
(6)	1	–	0	–	–
(7)	–	1	0	–	–
(3)(8)(9)(10)	–	–	–	–	1
(11)	1	1	1	–	–
(12)(13)	–	1	1	–	–

	E	D	C	B	A
(2)	–	–	–	1	–
(4)	0	1	–	–	–
(5)	0	–	1	–	–
(6)	1	–	0	–	–
(7)	–	1	0	–	–
(3)(8)(9)(10)	–	–	–	–	1
(11)(12)(13)	–	1	1	–	–

	E	D	C	B	A
(2)	–	–	–	1	–
(4)	0	1	–	–	–
(5)	0	–	1	–	–
(6)	1	–	0	–	–
(3)(8)(9)(10)	–	–	–	–	1
(7)(11)(12)(13)	–	1	–	–	–

	E	D	C	B	A
(2)	–	–	–	1	–
(5)	0	–	1	–	–
(6)	1	–	0	–	–
(3)(8)(9)(10)	–	–	–	–	1
(4)(7)(11)	–	1	–	–	–
(12)(13)					

Terms 3, 8, 9 and 10 are simplified by rule *3*.
Terms 12 and 13 are simplified by rule *1*.
Terms 11 and 12, and 13 are simplified by rule *3*.
Terms 7 and 11, 12, and 13 are simplified by rule *2*.
Terms 4 and 7, 11, 12, and 13 are simplified by rule *3*.

DUCTIONS IN THE BASIC PROCEDURES OF THE METHOD

t is possible to shorten the basic cedure which has been outlined in ail. This reduction is accomplished retaining the basic rules of com- ation and simplification and by adding following modifications.

. In order to record the original ction for simplification and transfor- tion:

The same procedure is followed as ined in items *1*, *2*, and *3* under "The ic Procedure of the Method" ex- t that the digital representation of each n instead of each product will be ten.

ple: $AB[CDE+A\bar{C}\bar{E}(C+D)]$

E	D	C	B	A		E	D	C	B	A	
(–	–	–	1	1)		[1	1	1	–	–	+
0	–	0	–	1		(–	1	1	–	–)]	

To record the function the digital esentation of each term is written, one n beneath the other. A *P* will be placed t to all terms that are products. An ill be placed next to all terms that are is.

Sample: To AND Form

E	D	C	B	A	
–	–	–	1	1	P
1	1	1	–	–	P
0	–	0	–	1	P
–	1	1	–	–	S

To OR Form

E	D	C	B	A	
–	–	–	1	1	P
1	1	1	–	–	P
0	–	0	–	1	P
–	1	1	–	–	S

2. Combination. The basic rules are applied for indicated combinations with these additions for transformation to the AND form: (*a*). A term that is a product will be combined as a whole term; (*b*). A term that is a sum (*S*) will be combined digit by digit.

For transformation to the OR form: (*a*). A term that is a sum will be combined as a whole term; (*b*). A term that is a product (*P*) will be combined digit by digit.

To illustrate the combination consider two products

	E	D	C	B	A	
$D\cdot B\cdot A=$	–	1	–	1	1	P
$\bar{E}\cdot D\cdot B=$	0	1	–	1	–	P

and two sums

	E	D	C	B	A	
$(\bar{E}+B+\bar{A})=$	0	–	–	1	0	S
$(\bar{D}+C+B)=$	–	0	1	1	–	S

Transformation to the sum of products (AND) form:

Example 1. Combination of two products

E	D	C	B	A		E	D	C	B	A
–	1	–	1	1	P	0	1	–	1	1
0	1	–	1	–	P					

Example 2. Combination of two sums

E	D	C	B	A		E	D	C	B	A
0	–	–	1	0	S	0	0	–	–	–
						0	–	1	–	–
–	0	1	1	–	S	0	–	–	1	–
						–	0	–	1	–
						–	–	1	1	–
						–	–	–	1	–
						–	0	–	–	0
						–	–	1	–	0
						–	–	–	1	0

Example 3. Combination of a product AND sum

E	D	C	B	A		E	D	C	B	A
–	1	–	1	1	P	0	1	–	1	1
0	–	–	1	0	S	–	1	–	1	1

Transformation to the product of sum (OR) form:

Example 4. Combination of two products

E	D	C	B	A		E	D	C	B	A
–	1	–	1	1	P	0	1	–	–	–
0	1	–	1	–	P	–	1	–	–	–
						0	–	–	1	–
						–	1	–	1	–
						–	–	–	1	–
						0	–	–	–	1
						–	1	–	–	1
						–	–	–	1	1

Example 5. Combination of two sums

E	D	C	B	A		E	D	C	B	A
0	–	–	1	0	S	0	0	1	1	0
–	0	1	1	–	S					

Example 6. Combination of a product OR a sum OR a product

E	D	C	B	A		E	D	C	B	A	
–	1	–	1	1	P	0	1	1	0	–	
0	–	1	0	–	S	0	–	1	0	1	
0	1	–	1	–	P	0	1	–	1	–	P

E	D	C	B	A
0	1	1	0	–
0	–	1	0	1
0	1	1	0	–
0	1	1	0	1

Upon close inspection of these examples it appears that some of the resulting terms of an indicated combination do not appear. This is due to part (*a*), item *1* under "Rules for Transforming the Function."

3. Simplification. The examples in the preceding section were not simplified. Further inspection of example *2* yields that one term contains only the *B* signal. By rule *3*, all other terms that contain *B* (a 1 in the *B* digit) are redundant and may be eliminated. This simplification may be accomplished at the same time the combination is made, thereby directly reducing the number of resulting terms. To accomplish this, the following rule may be used:

If in performing the indicated combination (for either transformation) of literals or whole terms, one of the items has its digital representation duplicated in the resulting term, all other combinations of this item will prove redundant and therefore need not be considered.

Example (from Example *2*)

	E	*D*	*C*	*B*	*A*	
	0	–	–	1	0	*S*
	–	0	1	1	–	*S*

E	*D*	*C*	*B*	*A*
–	–	–	1	–
0	0	–	–	–
0	–	1	–	–
–	0	–	–	0
–	–	1	–	0

Example (from Example *6*)

	E	*D*	*C*	*B*	*A*	
(2)	–	1	–	1	1	*P*
(3)	0	–	1	0	–	*S*
(4)	0	1	–	1	–	*P*

	E	*D*	*C*	*B*	*A*	
(2)(3)	0	1	1	0	–	
	0	–	1	0	1	
(4)	0	1	–	1	–	*P*

	E	*D*	*C*	*B*	*A*
(2)(3)(4)	0	1	1	0	–
	0	–	1	0	1

Sample:

To AND Form

(1)

	E	*D*	*C*	*B*	*A*	
(2)	–	–	–	1	1	*P*
(3)	1	1	1	–	–	*P*
(4)	0	–	0	–	1	*P*
(5)	–	1	1	–	–	*S*

(2)

	E	*D*	*C*	*B*	*A*	
	–	–	–	1	1	*P*
(3)	1	1	1	–	–	*P*
(4)(5)	0	1	0	–	1	

(3)

E	*D*	*C*	*B*	*A*
1	1	1	1	1
0	1	0	1	1

To OR Form

(1)

	E	*D*	*C*	*B*	*A*	
(2)	–	–	–	1	1	*P*
(3)	1	1	1	–	–	*P*
(4)	0	–	0	–	1	*P*
(5)	–	1	1	–	–	*S*

(2)

	E	*D*	*C*	*B*	*A*	
(2)	–	–	–	1	1	*P*
	1	–	0	–	–	
	1	–	–	–	1	
	0	1	–	–	–	
(3)(4)(5)	–	1	0	–	–	
	–	1	–	–	1	
	0	–	1	–	–	
	–	–	1	–	1	
	–	1	1	–	–	

(3)

		E	*D*	*C*	*B*	*A*
	a	–	–	–	–	1
	b	–	–	–	1	–
	c	1	–	0	–	–
	d	1	–	–	–	1
(2)(3)(4)(5)	*e*	0	1	–	–	–
	f	–	1	0	–	–
	g	–	1	–	–	1
	h	0	–	1	–	–
	i	–	–	1	–	1
	j	–	1	1	–	–

(4)

		E	*D*	*C*	*B*	*A*
	a	–	–	–	–	1
	b	–	–	–	1	–
	c	1	–	0	–	–
(2)(3)(4)(5)	*e*	0	1	–	–	–
	f	–	1	0	–	–
	h	0	–	1	–	–
	j	–	1	1	–	–

(5)

	E	*D*	*C*	*B*	*A*
a	–	–	–	–	1
b	–	–	–	1	–
c	1	–	0	–	–
e	0	1	–	–	–
h	0	–	1	–	–
f,j	–	1	–	–	–

(6)

	E	*D*	*C*	*B*	*A*
a	–	–	–	–	1
b	–	–	–	1	–
c	1	–	0	–	–
h	0	–	1	–	–
f,j	–	1	–	–	–

Obtaining the Minimum Form of a Solution

After all indicated combinations have been performed the resulting terms constitute a solution. This solution is not necessarily in the minimum form and the following process may be used to obtain a minimum form of a solution. The process will indicate that a term is redundant if its function is provided for in two other terms. In order for this to be valid the two terms must contain fewer or an equal amount of literals than the third term. The process is used as follows:

1. Consider the terms containing the greatest number of literals first.

(*a*). Substitute a 0 and a 1 in a digit containing a dash (–) in the considered term. This will yield two new forms of the considered term.

(*b*). Compare one of these new forms to the remaining terms. If it duplicates the literal digit representation of another term the other new form of the considered term will be compared. If this form duplicates the literal digit representation of another term the considered term is redundant and will be eliminated.

(*c*). Should the first 0,1 substitution not prove the term redundant the procedure will be repeated for another digit containing a dash in the considered term. This will continue until the term is determined redundant or that it is to be maintained.

2. The next term containing the greatest number of literals is then considered and checked for redundancy in the same manner. This is done until there are no more terms to give the necessary comparison, i.e., that there are not three terms that have at least an equal amount of literals.

The sample function for both transformations is already in the minimum form.

The following examples will illustrate the procedure.

Example 7:

	C	*B*	*A*
(2)	–	1	1
(3)	1	–	0
(4)	1	1	–

	C	*B*	*A*
(2)	–	1	1
(3)	1	–	0

All terms have the same number of literals so that each will be considered.

Term (*2*)

Making the 0,1 substitution in the digit yields the new forms of 0 1 1 and 1 1 1.

Comparing 0 1 1 to terms (3) and (4) shows that neither term has its literal digit representation duplicated. Therefore term (2) is not redundant and must be retained.

m (3)

'aking the 0,1 substitution in the B igit yields the new forms 1 0 0 and 1 0.

omparing the 1 1 0 form with terms) and (4) it duplicates the literal igit representation of term (4). The 1 0 form is then compared to term) and it does not duplicate this rm's literal digit representation. herefore term (3) is not redundant id must be retained.

m (4)

'aking the 0,1 substitution in the A igit yields the new forms 1 1 0 and 1 1.

omparing the 1 1 0 form to terms (2) id (3) it duplicates the literal digit presentation of term (3). The 1 1 1 rm is then compared to term (2) and duplicates the literal digit representa- on of term (2). Therefore term (4) redundant and is eliminated.

mple 8:

E	D	C	B	A
–	0	–	0	1
–	–	1	1	1
–	1	–	1	–
–	0	1	–	–
0	–	0	0	1
0	1	0	–	–

	E	D	C	B	A
(2)	–	0	–	0	1
(4)	–	1	–	1	–
(5)	–	0	1	–	–
(7)	0	1	0	–	–

ıbstitution of a 0, 1 in the D digit erm (6) yields the new forms which icate the literal digit representation erm (2) and (7). Therefore term is redundant and is eliminated.

Substitution of 0,1 in the D digit of term (3) yields new forms which duplicate the literal digit representation of terms (4) and (5). Therefore term (3) is redundant and is eliminated.

The other terms are checked for redundancies and none are found. Therefore all must be retained.

After becoming familiar with this process the substitution and comparisons will be made mentally and the redundant terms eliminated by inspection.

This process allows a term that is provided in two other terms to be eliminated. Elaboration of the process so that a function that is provided in three or more terms can be eliminated. Generally the elimination of these redundant terms will yield the minimum form of the solution. Many other methods (including those outlined in the Karnaugh, Quine, and McCluskey publications) may be used to obtain the minimum form of the solution.

Writing the Final Expression

After all indicated combinations have been performed and all simplifications have been made the remaining terms are in a group. For the conversion to the AND form each remaining term will be a product and the terms will be separated by OR functions. For the conversion to the OR form each remaining term will be a sum and the terms will be separated by AND functions.

For either conversion each term will be written as follows:

1. The variable will be present if a 1 appears in its digit of the particular term.
2. The negation of the variable will be written if a 0 appears in its digit of the particular term.
3. The variable will not be written for a particular term if a dash (–) appears in its digit.

Sample:

To AND Form

E	D	C	B	A	
1	1	1	1	1	$A\ B\ C\ D\ E$
0	1	0	1	1	$A\ B\ \bar{C}\ D\ \bar{E}$

Final expression: $X = ABCDE + AB\bar{C}D\bar{E}$

To OR Form

E	D	C	B	A	
–	–	–	1	–	B
0	–	1	–	–	$(C+\bar{E})$
1	–	0	–	–	$(\bar{C}+E)$
–	–	–	–	1	A
–	1	–	–	–	D

Final expression: $X = B(C+\bar{E})(\bar{C}+E)AD$
$= ABD(C+\bar{E})(\bar{C}+E)$

Figs. 2(A) and (B) show relay and static logic diagrams of the sample expression for both transformations.

Summary and Conclusions

In summation, two new methods of simplification and transformation of Boolean functions have been presented in this paper. These methods are utilized to obtain a minimum solution for a desired form and may be referred to as conversions of the original function. One method is used to convert the original function to a minimum sum of products and the other method is used to convert the original function to a minimum product of sums. The two methods parallel each other in that they utilize the same basic rules, the major difference being the recording of the original expression. Because of this similarity the two methods have been explained as one basic procedure. The fundamental principles of this basic procedure have been outlined in detail in the discussion. A reduction has also been explained that allows the basic procedure to be considerably shortened. A "sample" Boolean function and independent "examples" have been included to better explain the procedure. Examples of conversion to both forms of two Boolean functions have also been included in the discussion.

The following advantages occur through the use of the procedure presented in this paper:

1. The method for either conversion is based upon the postulates and theorems of Boolean algebra. Therefore, (a) any extension of the procedure or any portion of a

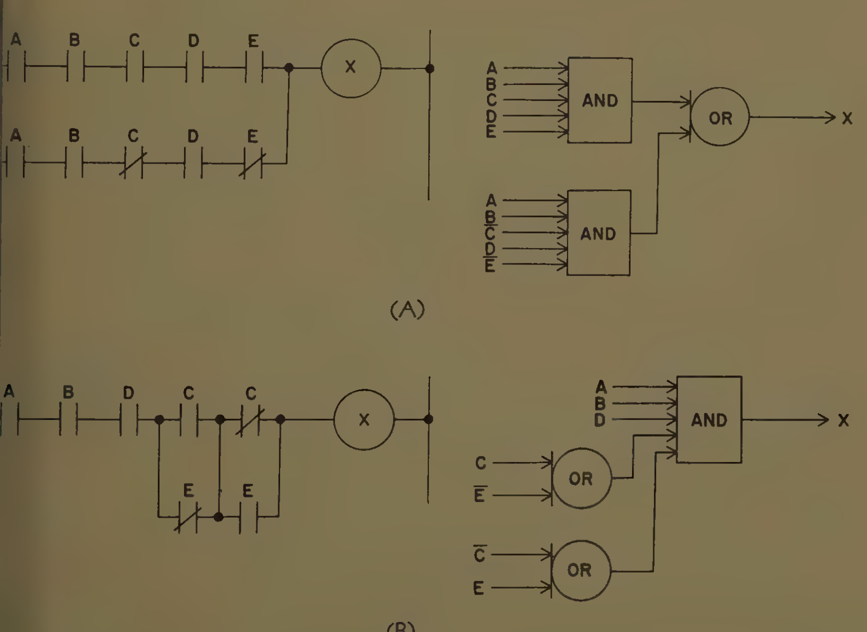


Fig. 2. Relay and static diagrams of transformation of sample expression

A—AND form
B—OR form

similar process that is also based on Boolean algebra may be incorporated to expedite the procedure; and (*b*) the manipulation of the original expression in lettered variable form by standard Boolean algebra methods is eliminated.

2. The original Boolean function appearing in either pure form (sum of products or product of sums) or a combination of both forms may be converted directly. Therefore, (*a*) it is not necessary that the original function be initially presented in either pure form to accomplish minimization or conversion, as it is in the Karnaugh map method; and (*b*) it is not necessary to analyze the original expression for the basic transmission functions as it is in McCluskey's method. This eliminates duplication of information stated initially in the original function. This duplication is encountered in the combining of these basic transmission functions to achieve the minimum form of a solution. For either conversion by the new methods the original function is the true base.

3. No restriction is placed on the number of variables or the complexity of the Boolean function for either conversion.

4. The basic procedure is the same for either conversion. Therefore by becoming acquainted with one of the new methods of conversion it can be applied easily to accomplish the other, eliminating the necessity of learning two distinctly different methods.

5. Although not a new concept the 1,0 and dash (–) variable representation utilized presents first, a clear, concise and easily handled representation of the original function for either conversion; second, a step-by-step tabular form for either conversion that will reduce the possibility of omitting literals of terms or whole terms; and third, a record that can be checked conveniently for errors.

6. The procedure is such that for either conversion most of the redundant terms are disregarded or eliminated as the conversion progresses. This means a minimum number of redundant terms are carried through to the last step.

In conclusion, from the fundamental precepts of the procedure presented:

1. New parallel methods of simplifying (obtaining a minimum form) Boolean functions have been provided for static and conventional relay-circuitry reduction. The equipment itself determines the conversion necessary to provide for maximum savings.

2. The use of standard Boolean algebra operation on the function has been eliminated in these new methods. Some of the restrictions encountered in using other methods (previously mentioned) have also been eliminated.

3. The rigidly systematic parallel methods and tabular 1,0 and dash (–) variable representation present an easily programmed process for digital computer application.

Extension of these basic ideas will surely provide even further reduction of the procedure.

Appendix

The preceding sample and examples considered no more than five variables. Th following examples illustrate how both transformations may be accomplished for expression containing more than five variables. The method utilized in these examples is the reductio of the basic procedure.

Example 1: $(C+\bar{D}+BF)[AB(\bar{C}+E)+\bar{B}D]+A\bar{B}C[(D+\bar{E}+F)+\bar{B}\bar{F}]$

To AND Form

F	*E*	*D*	*C*	*B*	*A*	
–	–	–	1	–	–	P
–	–	0	–	–	–	P
1	–	–	–	1	–	P
–	–	–	–	1	1	P
–	1	–	0	–	–	S
–	–	1	–	0	–	P
–	–	–	1	0	1	P
1	0	1	–	–	–	S
0	–	–	–	0	–	P

F	*E*	*D*	*C*	*B*	*A*
–	–	–	1	–	–
–	–	0	–	–	–
1	–	–	–	1	–
–	1	–	–	1	1
–	–	–	0	1	1
–	–	1	–	0	–
1	–	–	1	0	1
–	0	–	1	0	1
0	–	–	1	0	1
–	–	1	1	0	1

F	*E*	*D*	*C*	*B*	*A*
–	–	–	1	–	–
–	–	0	–	–	–
1	–	–	–	1	–
–	1	–	–	1	1
–	–	–	0	1	1
–	–	1	–	0	–
–	–	–	1	0	1
–	0	–	1	0	1
–	–	1	1	0	1

F	*E*	*D*	*C*	*B*	*A*
–	1	–	1	1	1
–	–	1	1	0	–
–	1	0	–	1	1
–	–	0	0	1	1
1	1	–	–	1	1
1	–	–	0	1	1
–	–	–	1	0	1

F	*E*	*D*	*C*	*B*	*A*
–	1	–	1	1	1
–	–	1	1	0	–
–	–	0	0	1	1
1	–	–	0	1	1
–	–	–	1	0	1

F	*E*	*D*	*C*	*B*	*A*	
–	1	–	1	–	1	ACE
–	–	1	1	0	–	$\bar{B}CD$
–	–	0	0	1	1	$AB\bar{C}\bar{D}$
1	–	–	0	1	1	$AB\bar{C}F$
–	–	–	1	0	1	$A\bar{B}C$

$$X=ACE+\bar{B}CD+AB\bar{C}\bar{D}+AB\bar{C}F+A\bar{B}C$$

To OR Form

F	*E*	*D*	*C*	*B*	*A*	
–	–	–	1	–	–	S
–	–	0	–	–	–	S
1	–	–	–	1	–	P
–	–	–	–	1	1	P
–	1	–	0	–	–	S
–	–	1	–	0	–	P
–	–	–	1	0	1	P
1	0	1	–	–	–	S
0	–	–	–	0	–	P

F	*E*	*D*	*C*	*B*	*A*
1	–	0	1	–	–
–	–	0	1	1	–
–	–	1	–	1	–
–	–	1	–	–	1
–	–	–	–	0	1
–	1	1	0	–	–
–	1	–	0	0	–
–	–	–	1	–	–
–	–	–	–	0	–
–	–	–	–	–	1
1	0	1	–	0	–

F	*E*	*D*	*C*	*B*	*A*
1	–	0	1	–	–
–	–	–	1	1	–
–	–	1	–	1	–
–	–	1	–	–	1
–	–	–	–	0	1
–	1	1	0	–	–
–	1	–	0	0	–
–	–	–	1	–	–
–	–	–	–	0	–
–	–	–	–	–	1

F	*E*	*D*	*C*	*B*	*A*
1	–	0	1	–	–
–	–	–	1	1	–
–	–	–	–	0	1
–	1	–	0	0	–
–	–	1	–	–	1
–	–	1	1	1	–
–	1	1	0	0	–
–	–	1	–	1	1
–	1	1	0	–	1

F	*E*	*D*	*C*	*B*	*A*	
1	–	0	1	–	–	$(C+\bar{D}+F)$
–	–	–	1	1	–	$(B+C)$
–	–	–	–	0	1	$(A+\bar{B})$
–	1	–	0	0	–	$(\bar{B}+\bar{C}+E)$
–	–	1	–	–	1	$(A+D)$

$$X=(C+\bar{D}+F)(B+C)(A+\bar{B})(\bar{B}+\bar{C}+E)(A+D)$$

Example 2: (The digits representing the variables will be reversed to those in the previo cases. This has been done deliberately to illustrate that the variables may be initial assigned to digits in any order. By assigning the corresponding variables to the corre digits throughout the procedure the order has no effect on the conversion.):

$$X=[AB\bar{C}(D+E+\bar{F})(C+\bar{G}+H)]D\bar{G}+[AGH+(\bar{A}+\bar{B}+F)(\bar{C}+\bar{D}+GH)]$$

To AND Form

(1)

A	B	C	D	E	F	G	H	
1	1	0	–	–	–	–	–	P
–	–	–	1	1	0	–	–	S
–	–	1	–	–	–	0	1	S
–	–	–	1	–	–	0	–	P
1	–	–	–	–	–	1	1	P
0	0	–	–	–	0	–	–	S
–	–	0	–	–	–	–	–	P
–	–	–	0	–	–	–	–	P
–	–	–	–	–	–	1	1	P

(2)

A	B	C	D	E	F	G	H	
1	1	0	1	–	–	0	–	
–	–	–	1	1	0	–	–	S
–	–	1	–	–	–	0	1	S
1	–	–	–	–	–	1	1	
0	–	0	–	–	–	–	–	
0	–	–	0	–	–	–	–	
0	–	–	–	–	–	1	1	
–	0	0	–	–	–	–	–	
–	0	–	0	–	–	–	–	
–	0	–	–	–	–	1	1	
–	–	0	–	–	0	–	–	
–	–	–	0	–	0	–	–	
–	–	–	–	–	0	1	1	

(3)

A	B	C	D	E	F	G	H	
1	1	0	1	–	–	0	–	
–	–	1	–	–	–	0	1	S
–	–	–	–	–	–	1	1	
0	–	0	–	–	–	–	–	
0	–	–	0	–	–	–	–	
–	0	0	–	–	–	–	–	
–	0	–	0	–	–	–	–	
–	–	0	–	–	0	–	–	
–	–	–	0	–	0	–	–	

(4)

A	B	C	D	E	F	G	H
1	1	0	1	–	–	0	–
–	–	–	–	–	–	1	1
0	–	0	–	–	–	–	–
0	–	–	0	–	–	–	–
–	0	0	–	–	–	–	–
–	0	–	0	–	–	–	–
–	–	0	–	–	0	–	–
–	–	–	0	–	0	–	–

(5)

A	B	C	D	E	F	G	H	
–	–	0	1	–	–	0	–	$\bar{C}D\bar{G}$
–	–	–	–	–	–	1	1	GH
0	–	0	–	–	–	–	–	$\bar{A}\bar{C}$
0	–	–	0	–	–	–	–	$\bar{A}\bar{D}$
–	0	0	–	–	–	–	–	$\bar{B}\bar{C}$
–	0	–	0	–	–	–	–	$\bar{B}\bar{D}$
–	–	0	–	–	0	–	–	$\bar{C}\bar{F}$
–	–	–	0	–	0	–	–	$\bar{D}\bar{F}$

$$X=\bar{C}D\bar{G}+GH+\bar{A}\bar{C}+\bar{A}\bar{D}+\bar{B}\bar{C}+\bar{B}\bar{D}+\bar{C}\bar{F}+\bar{D}\bar{F}$$

To OR Form

(1)

A	B	C	D	E	F	G	H	
1	1	0	–	–	–	–	–	P
–	–	–	1	1	0	–	–	S
–	–	1	–	–	–	0	1	S
–	–	–	1	–	–	0	–	P
1	–	–	–	–	–	1	1	P
0	0	–	–	–	0	–	–	S
–	–	0	–	–	–	–	–	S
–	–	–	0	–	–	–	–	S
–	–	–	–	–	–	1	1	P

(2)

A	B	C	D	E	F	G	H	
1	1	0	–	–	–	–	–	P
–	–	–	1	1	0	–	–	S
–	–	1	–	–	–	0	1	S
–	–	–	1	–	–	0	–	P
1	–	–	–	–	–	1	1	P
0	0	–	–	–	0	–	–	S
–	–	0	0	–	–	1	–	
–	–	0	0	–	–	–	1	

(3)

A	B	C	D	E	F	G	H	
1	1	0	–	–	–	–	–	P
–	–	–	1	1	0	–	–	S
–	–	1	–	–	–	0	1	S
–	–	–	1	–	–	0	–	P
–	–	0	0	–	–	1	–	
–	–	0	0	–	–	–	1	
0	0	–	–	–	0	1	–	
0	0	–	–	–	0	–	1	

(4)

A	B	C	D	E	F	G	H
–	–	0	0	–	–	1	–
–	–	0	0	–	–	–	1
0	0	0	–	–	0	1	–
0	0	0	–	–	0	–	1
0	0	–	1	1	0	1	–
0	0	–	1	1	0	–	1
0	0	1	–	–	0	0	1
0	0	–	1	–	0	1	–
0	0	–	1	–	0	–	1
0	0	–	–	–	0	0	1

(5)

A	B	C	D	E	F	G	H	
–	–	0	0	–	–	1	–	$(\bar{C}+\bar{D}+G)$
–	–	0	0	–	–	–	1	$(\bar{C}+\bar{D}+H)$
0	0	–	1	–	0	1	–	$(\bar{A}+\bar{B}+D+\bar{F}+G)$
0	0	–	–	–	0	0	1	$(\bar{A}+\bar{B}+\bar{F}+\bar{G}+H)$

$$X=(\bar{C}+\bar{D}+G)(\bar{C}+\bar{D}+H)(\bar{A}+\bar{B}+D+\bar{F}+G)(\bar{A}+\bar{B}+\bar{F}+\bar{G}+H)$$

References

1. The Map Method for Synthesis of Combinational Logic Circuits, M. Karnaugh. *AIEE Transactions*, vol. 72 (*Communication and Electronics*), pt. I, Nov. 1953, pp. 593–99.

2. The Problem of Simplifying Truth-Functions, W. V. Quine. *The American Mathematical Monthly*, Menasha, Wis., vol. 59, Oct. 1952.

3. Minimization of Boolean Functions, E. J. McCluskey, Jr. *Bell System Technical Journal*, New York, N. Y., vol. 35, Nov. 1956, pp. 1417–44.

Discussion

Peter L. Randlev (IBM Research, Yorktown Heights, N. Y.): While Mr. Howard has been correct in attempting to find a simple straightforward method to obtain the most economical switching circuit, his method does not consistently do this.

The most economical form of a combinational circuit is dependent upon the components to be utilized. However, in developing a method for determining the configuration of such a circuit, it is not sufficient just to specify whether an alternational (sum of products) or a conjunctive (product of sums) form of logic shall be used. Often, the most economical configuration should be a combination of both these forms. It is also important to make provision for additional rules to obtain this configuration in view of: 1. the cost of each term in an expression as well as the cost of each literal, 2. the relative cost of terms composed of various quantities of literals when, as is most often the case, this cost is not linear with the quantity of literals, and 3. the cost of a circuit configuration composed of several levels of logic as would be obtained if the function were in a factored rather than a normal (pure) form.

Any method that is to be able to find the most economical circuit configuration must be able to take into account factors such as these for defining a minimum. The author has no definition of a minimum.

For any method to be easily and successfully programmed for a digital computer application, as Mr. Howard suggests, it must be both systematic and unambiguous. As presented, his method fails this criterion. The rules are not stated explicitly enough and, rather than being rigidly systematic, the examples themselves are demonstrations that the rules allow any of several sequences in performing the operations. To be used with a digital computer, it would obviously be necessary to use some means other than lines of several lengths to indicate the sequence of groupings required in transformation. This system does not even appear to be the best with hand methods as the comparison of the relative lengths of the connecting lines would become quite difficult to see in large problems. However, the author should be able to modify quite simply his rules and notation for use with a digital computer.

Other than the purely clerical operations of recording the original function and writing the final expression, the method may be divided into three general operations: 1. transformation or combination, 2. simplification, and 3. minimization.

The transformation operation changes the original function into a normal (pure) form. As desired, this would be either the alternational normal form or the conjunctive normal form. The simplification operation apparently is intended to find the terms that have the fewest possible literals. This is accomplished by pairing the given terms and removing those terms that are found to be completely contained within one other term. However, the rules do not always allow the term with the fewest literals to be found. The following is such an example:

Example A:

Original data	Desired answer
– 0 1 0	– 0 1 0
– 1 0 1	– 1 0 1
1 – 1 1	1 – – –
1 – 0 0	
1 1 1 –	
1 0 0 –	

The stated rules will not allow any simplification of this original function.

It is probably Mr. Howard's intention to define minimization as that expression of the function whose sum of the total quantity of literals in all terms plus the quantity of terms is the smallest possible. As the author has apparently recognized (in the last paragraph under "Obtaining the Minimum Form of a Solution"), his method does not necessarily find a minimum solution under this definition. The operation the author has described for use in minimization is to remove redundant terms; that is, terms that are expressed by two other terms. This method often removes a term that should appear in the final answer. For example, see the following:

Example B:

Original data	Desired answer
1 0 1 – –	1 0 1 – –
– 1 1 0 1	– 1 1 0 1
1 1 – 0 0	1 1 – 0 0
0 – 1 1 0	0 – 1 1 0
0 0 – 1 1	0 0 – 1 1
1 – 1 0 –	
– 0 1 1 –	

If the rules are followed consistently, the first term of the original data will be removed by the last two terms which will themselves remain intact. This nonminimal answer would be

– 1 1 0 1
1 1 – 0 0
0 – 1 1 0
0 0 – 1 1
1 – 1 0 –
– 0 1 1 –

In other cases these same rules fail to remove redundant terms, as in the following

Example C:

Original data	Desired answer
1 – 1 1	1 – 1 1
1 – 0 0	1 – 0 0
– 1 1 0	– 1 1 0
– 1 0 1	– 1 0 1
1 1 – –	

When the rules are more precisely stated as to the sequence of operations for use with a digital computer, the sequence of the original data becomes important in attempting to find a minimum solution. This is demonstrated by example *D*. If this problem is "minimized" with the data both in its original and in an inverted sequence, two answers of different cost are obtained.

Example D:

Original Data	Desired answer	Nonminimal "minimized" answer
1 1 – 1	1 1 – 1	1 0 1 –
1 0 1 –	1 0 1 –	0 1 0 –
0 1 0 –	0 1 0 –	1 – 1 1
1 – 1 1		– 1 0 1
– 1 0 1		

In the reference 2 cited by the author, Professor Quine discusses the minimization of problems he calls "irredundant." Professor Quine's example is

Example E:

Original data	Desired answers		
1 0 –	1 0 –		0 1 –
0 1 –	0 – 1	or	1 – 0
– 1 0	– 1 0		– 0 1
– 0 1			

As Mr. Howard's method is completely dependent on finding redundant terms, it cannot find the minimum solution from these irredundant functions. The desired answers of these five examples can all be checked by other means (such as a truth table) to insure that they are equivalent to the original data. These problems are only samples of those problems that cannot be minimized by the author's rules.

It is my experience that as the problems become larger and more complex, finding a problem belonging to a class which cannot be minimized by this method becomes a more frequent occurrence. For a real minimization process to be useful, it must be possible to recognize the nonminimal cases even if a minimum answer cannot always be found (so that they may be solved by other methods).

There is no step in the author's method in which the intermediate results become more complex in form than the original. Thus, rather than referring to this as a minimization process, it would be more acceptable to call this portion of the author's paper a method to remove redundancy.

Extensions of the method to allow it first, to recognize cases where it cannot find a minimum, or second, to find a minimum answer for all cases, will make the process more complex and lengthy rather than reduce the necessary procedure. It is also necessary that a minimization procedure include the ability to handle "don't-care" conditions. These are normally created by specific input combinations that can't exist and taking them into consideration often allows drastic reductions in the required circuitry.

This method becomes, in common with all other methods known to this discusser, extremely difficult to use by hand for problems much larger than those that can be simplified by inspection, just because the problems are large. However, the large-problem area is the one in which the greatest savings may be made by the use of simplification schemes and which, therefore, requires the most work.

W. Stuart (Industrial Nucleonics Cor- ration, Columbus, Ohio): Mr. Howard s done a thorough job in presenting a thod of converting and simplifying nbinational logic circuits. Circuit de- n based upon the use of logic functions s received increasing attention in recent irs.

The binary nature of the normal industrial ntrol provides a comparatively simple ical system. It can be shown that the s of "symbolic logic" are applicable in a her restricted sense. Specifically, de- es capable of providing the logical operties of assertion and negation are basic requirement. Assertion corre- onds with ON while negation is representa- e of OFF.

Assertion can be categorized into two ncipal groups. Alternative assertion is scribed by the verbalization OR whereas ordinate assertion (coincidence) is de- iptive of AND. Negation is by definition overt representation of the absence of sertion. This inversion is often termed complementary function. The prac- al importance of grasping the concept negation cannot be overemphasized. gation is not a passive zero state. For ample, a normally closed contact does t provide negation unless it is also con- cted to a voltage source. Furthermore unoperated normally open contact is t synonymous with negation, only com- tible with it. An active signal must be ovided to represent the open condition this normally open contact in its un- erated state. Identical reasoning applies static switching devices.

It is possible and practical to construct ire logical systems, including memory, lizing only negation and one of the two sertions. Therefore, a given system ght contain only OR and NOT. An ernative system could contain only D and NOT. It should be evident that system lacking negation, or NOT, is not ssible.

A system composed of all three basic ic functions is defined as an English ic type. It provides the most flexible stem design choice and is apparently the sest approximation of normal thought evidenced by the language structure of least the English-speaking people. From ractical standpoint, such a logical system easier to understand, design, and trouble- oot. This consideration is important ce the basic description of the operation quence is invariably presented in a rbalized form. Furthermore, the usual ntrol design methods depend on this xibility unless a mathematical tool such the algebra of logic (Boolean algebra) utilized. Thorough mastery of this tool necessary to allow effective design. is algebra is limited in over-all applica- n since time variant or sequential opera- n is difficult to take into account. There- e, the entire class of memory types is cluded. However, for optimizing ex- mely large systems or very complex ings of combinational logic configura- ns, this algebra is very effective and rth while.

The usual starting place in tackling a ntrol design is with a formulation of the nditions necessary to activate each tput. Upon completion of the individual cuits as well as after final integration of the over-all system, it is desirable to expend a reasonable effort in checking and simplification. The art of simplification is tied closely to the economics of alternative logic components available. In the case of the static-switching system under consideration, the most economical system is generally achieved by the maximum use of OR units. Furthermore, the manner in which memory units are formed has an important bearing upon formation of optimum circuit design. Broadly, two processes can be used to affect simplification of combination circuits composed of ORS, ANDS, and NOTS or variations thereof. These processes are conversion and inversion. Conversion is concerned principally with OR–AND circuits, while inversion is concerned with NOT functions as the term implies.

A given portion of a combinational circuit may be the termination of two or more ANDS at a single OR. Conversely, it may be the termination of two or more ORS at a single AND. It can be shown that a one-to-one correspondence exists between these alternative circuits such that one may be converted to the other with about an even chance that simplification will result.

The greatest possibility for worth-while simplification exists whenever a given combinational circuit contains a repetition of one or more identical signals. The final circuit may or may not contain repetitive signals since economy is not primarily a logical consideration. It is equally important to note that if a signal A and its negation or complement $\bar{A}$ when found in a given combination are often but not invariably redundant.

Frequently, the availability of the normally closed contacts of sensing devices permits simplification by permitting inversion of the circuit. Inversion entails changing OR conditions to ANDS and AND conditions to ORS with all input signals inverted (negated) and providing the complement, or NOT, of the resultant circuit.

Both conversion and inversion may be carried out on a given circuit for the ultimate in simplicity. The type of memory circuit with which the combinational circuit is connected in a large measure determines whether conversion, inversion or both are profitable.

This paper presents a complete and usable method of handling OR–AND conversions and performing a straightforward procedure of simplification.

Memory circuits would appear to be most effectively handled with sequence or time-bar charts to define the combinational functions which occur momentarily that must be remembered for subsequent use in the circuit operation. Unless such a method is used to reduce the theoretical memory possibilities to actual requirements, the task is almost unmanageable. At any rate I have not seen a really practical method of handling memory functions with the use of Boolean algebra alone. The analysis of memory situations certainly should present a fruitful field for future investigation.

R. L. Howard: The author wishes to express his gratitude to Mr. Stuart and Mr. Randlev for their discussions.

The author's background is basically that of industrial-control systems while Mr. Randlev's is predominantly concerned with digital computer applications. His discussion therefore provides an appraisal of the procedure from the computer viewpoint, which is an area with which the author is least acquainted. Primarily Mr. Randlev's comments are directed to certain generalities contained in the paper that would seem to imply that the scope of the procedure presented is unlimited. It was not the author's intention to create this false impression, particularly regarding the minimization of Boolean functions.

The procedure presented does not provide methods for obtaining the most economic solution for a particular application or function. Surely many factors must be taken into consideration before the optimum form can be determined and it has been assumed that the designer has solved this portion of the problem before utilizing the procedure. Those applications that do find their most economic solution in one of the pure logic forms can be resolved directly from the procedure. Other applications that find their optimum solution in the combinational form, will require additional effort on the part of the designer in using the procedure as is the case with all similar methods known to the author.

Different lengths of lines are utilized throughout the procedure to indicate groupings within the function as well as serving as operational devices. This does not mean that the procedure is restricted to these symbols any more than the AND or OR functions in Boolean algebra are always represented by the same symbol. The author readily agrees that other designations can be utilized for the desired indication and that these may allow easier handling of the procedure.

The "don't-care" type of term has not been included in the procedure and what would be involved in taking this into consideration can only be determined by further investigation. In the same vein, it is the author's contention that only further examination can determine the length of the procedure necessary to enable it to derive an absolute minimum for a particular function.

It is a well-known fact that vast amounts of study have been directed toward the area of minimizing Boolean functions. Only a fraction of this effort could be covered in a paper of this length. Being aware of this, I made an attempt to qualify the rules for minimization, but obviously this qualification was not stressed to a great enough degree. Mr. Randlev's comments on the simplification and minimization portion of the paper have pinpointed the restriction of the rules covering these sections and have brought these limitations firmly to the attention of the author and the reader. These comments (including the author's presumed definition of minimization) are accurate and are invaluable to the author and the paper.

In summarizing Mr. Randlev's discussion, I do not feel that the rules presented in the paper have been shown to be incorrect but rather that they do not include all cases.

Mr. Randlev's and Mr. Stuart's discussions have been of great assistance. It is constructive analysis such as this that will increase the value of the procedure as a tool in logic design.

A Stability Criterion for Nonlinear Systems

Y. H. KU
FELLOW AIEE

A. A. WOLF
MEMBER AIEE

IN ANOTHER PAPER[1] Wolf reports on the recurrence relations in the solution of a certain class of nonlinear systems. In Wolf's dissertation[2] a mathematical theory for the analysis of a class of nonlinear systems has been developed. This paper reports on a stability criterion based on the aforementioned dissertation and its application to nonlinear as well as linear feedback control systems.

The method of stability analysis consists essentially of first finding the response of a nonlinear system to a given forcing function by either the method of partition given in reference 1 or the Taylor-Cauchy transform method.[3] These methods give the solution of certain class of nonlinear systems[1,2] in the form of a power series, such that the response $x(t)$ is governed by a function of the exponential type given in the following equation.

$$x(t)=\sum_{n=0}^{\infty} h_n t^n \qquad (1)$$

where h_n are constants for different values of n.

The Laplace transform of $x(t)$ is

$$X(s)=\sum_{n=0}^{\infty} \frac{h_n n!}{s^{n+1}} \qquad (2)$$

where s is a complex variable.

Stability Criterion

The stability criterion is given as follows. Given $x(t)$ the solution of a nonlinear differential equation (which represents a given nonlinear physical system) such that $x(t)=0$ $(\epsilon^{\sigma t})$. Then $x(t)\rightarrow 0$ as $t\rightarrow\infty$ if and only if the convex singularity hull (CSH) of $X(s)$ is entirely in the left-half plane (LHP).

The CSH is constructed by joining the points of singularity of the transform $X(s)$ in the s-plane. It may be noted that the singularities of $X(s)$ are not necessarily poles.

The nonlinear differential equation can be given in the following form

$$Z(D)x(t)+F(x, x', \ldots)=r(t) \qquad (3)$$

where

$Z(D)$ = linear differential operator
$F(x, x', \ldots)$ = nonlinear function of x, its derivative x', ...
$r(t)$ = the forcing function or input
$x(t)=c(t)$ = the response
$D=d/dt$ = the differential operator
$D^2=d^2/dt^2$, ...
$x'=dx/dt$ = the derivative of the response with respect to time t

A nonlinear system is said to be stable if to every bounded-decaying driving function or input and for all initial conditions the response $x(t)$ approaches zero as time increases to infinity. If the foregoing condition is not satisfied, the system may be oscillating in the steady-state or may run away and become unstable.

In a linear system no mention is usually made of bounded values when speaking of the stability of the system. The reason for this is that a linear system has a response whose form is essentially independent of the initial conditions. In a nonlinear system, it may go from a stable to an unstable state during some transition, such as switching which may alter the initial conditions. In such cases the nonlinear system is considered conditionally stable and the range of input functions and initial conditions must be specified for stable operation.

In linear systems the CSH is composed of poles for lumped parameter systems. For stability these poles must be confined to the LHP. In nonlinear systems the CSH is composed of singularities (which are not necessarily poles) and for stability these singularities must be confined to the LHP. Thus the stability criterion given previously holds for nonlinear as well as linear systems.

Construction of the CSH

The CSH and other related concepts in this paper are obtained from entire function theory.[4] For the purpose of this paper, however, the following simple examples show how CSH can be obtained for stability studies of physical systems.

Example 1

$$X(s)=\sum_{n=0}^{\infty} \frac{1}{s^{n+1}} \qquad (4)$$

This is a special form of equation 2, with $h_n n!=1$. Equation 4 or 2 (in general) looks like a Laurent series. However, one must not conclude that at $s=0$, $X(s)$ has an essential singularity. Instead it may be that $X(s)$ at $s=0$ is a regular point.

Clearly equation 4 is an expansion of

$$X(s)=\frac{1}{s-1} \qquad (5)$$

At $s=0$, $X(s)=-1$, while at $s=1$, $X(s)$ has a pole. So the series of equation 4 converges outside of the circle of convergence whose radius is $R=1$. In general, if the power series of equation 2 converges outside of the circle of convergence whose radius is $R=a$, one must not conclude that at $s=0$, $X(s)$ has an essential singularity.

Since the singularity $s=1$ is not in LHP, the system is unstable. This can be verified by taking the inverse transform of equation 5.

Example 2

$$X(s)=\sum_{n=0}^{\infty} \frac{a^n}{s^{n+1}} \qquad (6)$$

Compared with equation 2, $h_n n!=a^n$, where a is a constant. Equation 6 is an expansion of

$$X(s)=\frac{1}{s-a} \qquad (7)$$

At $s=a$, $X(s)$ has a singularity, which is a pole. The system is unstable.

Example 3

$$X(s)=\sum_{n=0}^{\infty} \frac{(-1)^n a^n}{s^{n+1}} \qquad (8)$$

is an expansion of

$$X(s)=\frac{1}{s+a} \qquad (9)$$

At $s=-a$, $X(s)$ has a pole, which is in LHP. Hence the system is stable.

Example 4

Let the simplified from of a given power series be given by

Paper 59-231, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted October 15, 1958; made available for printing December 15, 1958.

Y. H. Ku is with the Moore School of Engineering, University of Pennsylvania, Philadelphia, Pa.; A. A. Wolf, formerly with the Moore School of Engineering, is now with the Stromberg Carlson Company, Division of General Dynamics, Rochester, N. Y.

The authors wish to acknowledge the support given by the National Science Foundation for their research on nonlinear analysis.

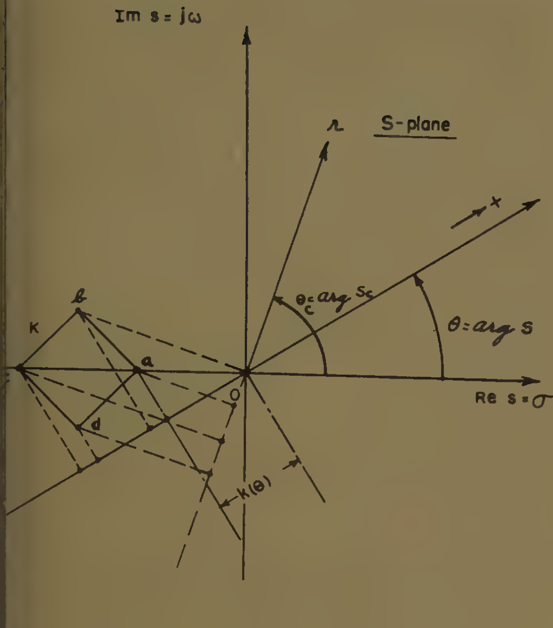
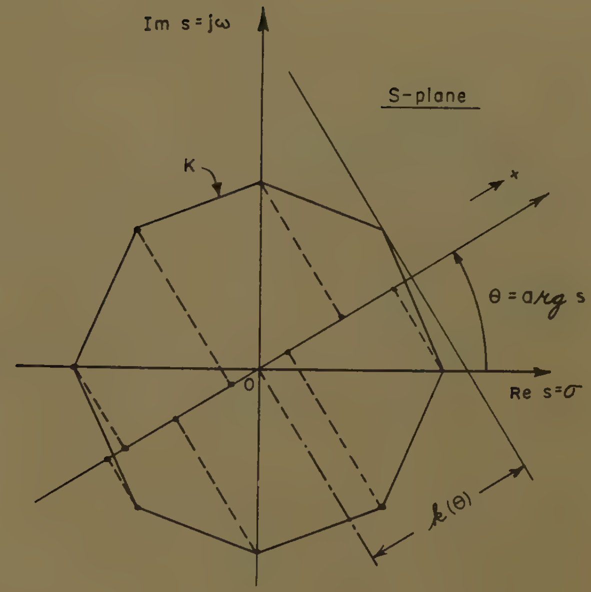


Fig. 1 (left). CSH for a stable system

Fig. 2 (right). CSH for an unstable system. Set of points K and its projection on $\theta = \arg s$



$$s) = \frac{1}{s+2} + \frac{1}{s+1-j2} + \frac{1}{s+1+j2} \qquad (10)$$

e singularities are at $s=-2$, $s=-1+$ and $s=-1-j2$. Connecting all three nts gives the CSH in the form of a tri-gle. Since the CSH lies entirely in the P, the system is stable.

The foregoing simple examples are en from the linear systems so that the struction of the CSH can be demon-ated in a straightforward manner. . 1 shows a CSH for a stable system, ile Fig. 2 shows a CSH for an unstable tem.

dius of Convergence and Direction of Singularity

From equations 4, 6, and 8 in examples , and 3 of the foregoing, it is seen that origin in the s-plane is not included as ngularity. In example 1, the circle of vergence has a radius $R=1$. In ex-ples 2 and 3, the circle of convergence a radius $R=a$. It is noted that in mples 1, 2, and 3 the singularity lies on circle of convergence. In general, the ius of the circle of convergence is given

$$\overline{\lim_{n\to\infty}}\, |n!\,h_n|^{1/n} \qquad (11)$$

xample 1, equation 11 gives $R=1$. In mples 2 and 3 equation 11 gives

$$\overline{\lim_{n\to\infty}}\, |n!\,h_n|^{1/n} = \overline{\lim_{n\to\infty}}\, |a^n|^{1/n} = a \qquad (12)$$

direction of the singularity (which lies he circle of convergence) is denoted by angle θ, which can be obtained as ows. Let

$$\sum_{n=0}^{k} \binom{k}{n} \epsilon^{jn\theta} \beta_{n+1} \qquad (13)$$

where

$$\beta_{n+1} = \frac{n!\,h_n}{R^{n+1}}$$

and

$$\binom{k}{n} = \frac{k!}{n!(k-n)!} \qquad (14)$$

Then θ is the direction of the singularity with radius R if

$$\overline{\lim_{k\to\infty}}\, |d_k|^{1/k} \geq 2 \qquad (15)$$

With the radius of the circle of convergence obtained from equation 11 and with the direction of singularity determined by equations 13 and 15, the singular point lies at

$$s = R\epsilon^{j\theta} \qquad (16)$$

The singularity given by equation 16 is then subtracted from $X(s)$ given by equation 2 and the process is repeated until all singularities are removed.

From example 1, $\beta_{n+1}=1$, and

$$d_k = \sum_{n=0}^{k} \binom{k}{n} \epsilon^{jn\theta} \qquad (17)$$

Then

$$\overline{\lim_{k\to\infty}} \left| \sum_{n=0}^{k} \binom{k}{n} \epsilon^{jn\theta} \right|^{1/k} = \overline{\lim_{k\to\infty}} \left| (1+\epsilon^{j\theta}) \right|^{1/k} = |1+\epsilon^{j\theta}| \qquad (18)$$

It is seen that the right-hand member of equation 18 is equal to 2 for $\theta=0$.

Application to a Nonlinear Feedback System

Consider the nonlinear feedback system of Fig. 3. The differential equation governing the dynamics of the system is given in the following

$$\frac{d^2x}{dt^2} + 2\frac{dx}{dt} + x + x^2 = r(t) \qquad (19)$$

The forcing function or input as shown Fig. 4 and given by

$$r(t) = \epsilon^{-4t} - 2\epsilon^{-3t} \qquad (20)$$

Assume that the initial conditions are: $x(0)=0$ and $x'(0)=1$. Substituting equation 20 in equation 19 and partitioning at the first linear term gives

$$D^2x = \frac{d^2x}{dt^2} = \epsilon^{-4t} - 2\epsilon^{-3t} - (2x' + x + x^2) \qquad (21)$$

The auxiliary equation is[1,2]

$$\sum_{n=0}^{\infty} C_n t^n = \epsilon^{-4t} - 2\epsilon^{-3t} - (2x' + x + x^2) \qquad (22)$$

and the solution of equation 21 is given by[1,2]

$$x(t) = \sum_{n=0}^{\infty} C_n Q_n(t) + K_0 + K_1 t \qquad (23)$$

where

$$Q_n(t) = \int_0^t \tau^n y(t-\tau)\,d\tau \qquad (24)$$

and

$$y(t) = \frac{1}{2\pi j}\int_{Br} \frac{\epsilon^{st}}{s^2}\,ds = tu(t) \qquad (25)$$

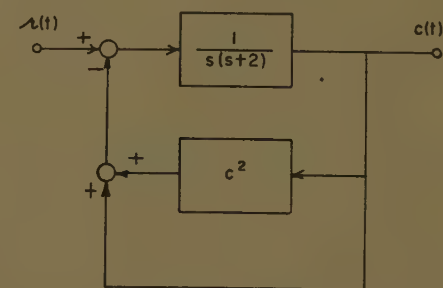


Fig. 3. Nonlinear feedback system

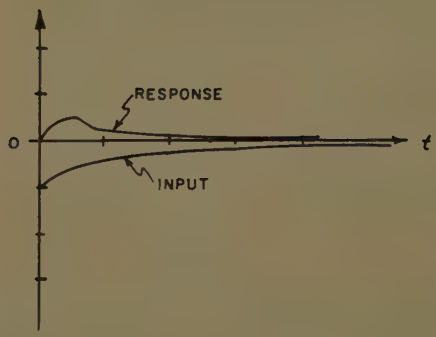


Fig. 4. Sketch of input and response functions

where Br is the Wagner-Bromwich contour enclosing the origin and $u(t)$ is the unit step function. Note that s^2 inside the integral is a special form of $Z(s)$, which depends upon the form of $Z(D)$ in equation 3. In equation 21, $Z(D)=D^2$; hence $Z(s)=s^2$. Substituting equation 25 in equation 24 gives

$$Q_n(t)=\int_0^t \tau^n(t-\tau)d\tau=\frac{t^{n+2}}{(n+1)(n+2)} \tag{26}$$

Applying the initial conditions gives $K_0=0$ and $K_1=1$ in equation 23. Thus

$$x(t)=\sum_{n=0}^{\infty} C_n \frac{t^{n+2}}{(n+1)(n+2)}+t \tag{27}$$

Substituting eqution 27 in equation 22 and solving for the constants recursively[1] gives

$$C_0=-3,\ C_1=7,\ C_2=-15/2,\ C_3=31/6,\ \ldots \tag{28}$$

Substituting these values in equation 27 gives

$$x(t)=t-\frac{3}{2}t^2+\frac{7}{6}t^3-\frac{15}{24}t^4+\frac{31}{120}t^5-\ldots \tag{29}$$

The n th term is recognized as

$$\frac{(-1)^{n+1}(2^n-1)}{n!}t^n$$

so that

$$x(t)=\sum_{n=0}^{\infty}\frac{(-1)^{n+1}(2-1)}{n!}t^n \tag{30}$$

The Laplace transform of equation 30 is

$$X(s)=\sum_{n=0}^{\infty}\frac{(-1)^{n+1}(2^n-1)}{s^{n+1}} \tag{31}$$

Stability Study of the Feedback System

From equation 11, the modulus of the first singularity is

$$R_1=\overline{\lim_{n\to\infty}}\left|(-1)^{n+1}(2^n-1)\right|^{1/n}=2 \tag{32}$$

Since the radius is not unity, application of equation 14 gives

$$\beta_{n+1}=-(-1)^n(2^n-1)2^{-n-1} \tag{33}$$

Substituting equation 33 for β_{n+1} in equation 13 gives

$$d_k=-\sum_{n=0}^{k}\binom{k}{n}(-1)^n(2^n-1)2^{-n-1}\epsilon^{jn\theta} \tag{34}$$

It can be shown that for $\theta=\pi$

$$\overline{\lim_{k\to\infty}}\left|\sum_{n=0}^{k}\binom{k}{n}2^{-n-1}(2^n-1)\right|^{1/k}=2 \tag{35}$$

Thus the singularity is at

$$s_1=2\epsilon^{j\pi} \tag{36}$$

If $\beta_{n+1}R^{n+1}$ is a power function s_1 is a pole.

From example 3, equation 8 gives

$$X_1(s)=\frac{1}{s+2}=\sum_{n=0}^{\infty}\frac{(-1)^n2^n}{s^{n+1}} \tag{37}$$

for $a=2$. Subtracting $X_1(s)$ from $X(s)$ as given by equation 31 yields

$$X_2(s)=X(s)-X_1(s)=-\sum_{n=0}^{\infty}\frac{(-1)^n}{s^{n+1}}. \tag{38}$$

The singularity of equation 38 is at

$$s_2=\epsilon^{j\pi} \tag{39}$$

by inspection. Plotting the singularity hull, shown in Fig. 5, and applying the stability criterion, the system is stable.

In this example, the singularities are poles. In simplified form, equation 31 corresponds to

$$X(s)=\frac{1}{s+1}-\frac{1}{s+2} \tag{40}$$

Taking the inverse transform of equation 40 gives

$$x(t)=\epsilon^{-t}-\epsilon^{-2t} \tag{41}$$

which is the solution of equation 21 in closed form. The first derivative is

$$x'(t)=-\epsilon^{-t}+2\epsilon^{-2t} \tag{42}$$

At $t=0$, equations 41 and 42 give respectively $x(0)=0$ and $x'(0)=1$.

Conjugate Indicator Function and Diagram

So far the tranform pairs $x(t)$ and $X(s)$ have been considered. Suppose that t be a complex variable denoted by

$$t=r\epsilon^{j\theta} \tag{43}$$

where $r=|t|$. Consider Fig. 6. The growth of $x(t)$ is determined by the indicator function

$$f(\theta)=\overline{\lim_{r\to\infty}}\frac{\log|x(t)|}{r} \tag{44}$$

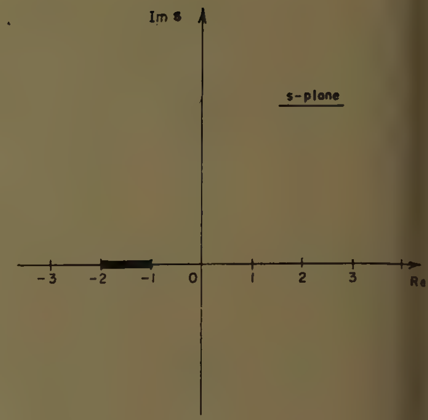


Fig. 5. Singularity hull of X(s)

This diagram involving θ and t is called th conjugate indicator diagram. The fun tion $f(\theta)$, which is also called the Phrag mén-Lindelöf function, determines th different behaviors of $x(t)$ in the differen directions of t as $r\to\infty$. To understan the meaning of $f(\theta)$ consider

$$x(t)=\epsilon^{(a+jb)t} \quad (4$$

Substituting equation 43 in equation and simplifying gives

$$|x(t)|=\epsilon^{r(a\cos\theta-b\sin\theta)} \quad (4$$

The behavior of $|x(t)|$ depends first of a on $|t|$, which is independent of θ. The as $t\to\infty$ the growth of $x(t)$ depends on th sign of the factor involving θ. That is say, along different rays of t (as define by equation 43), the value of θ determin whether ($a\cos\theta-b\sin\theta$) is negative positive. If it is negative it is clear th $|x(t)|\to 0$ as $r\to\infty$. It is also clear th there is a range of θ for which the fact ($a\cos\theta-b\sin\theta$) is positive. In the cases $|x(t)|$ will be unbounded as r ten to infinity. The function $f(\theta)$ is thus measure of whether $|x(t)|$ is stable or no It may be noted here that the properti of $f(\theta)$ are the basis for the developme of the stability criterion given in the for going.

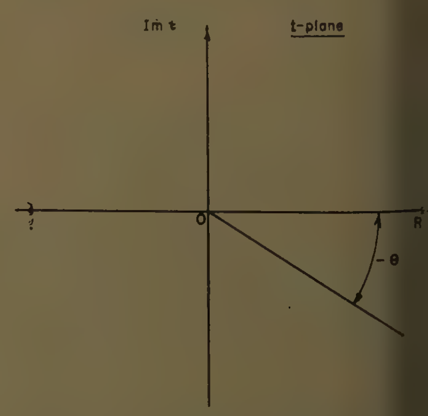


Fig. 6. Conjugate indicator diagram

nce we deal with positive real time, nly need consider results of $f(\theta)$ for in Fig. 6 if we are interested in the lity of a system but not the degree ability. It is possible to utilize the mation obtained from $f(\theta)\neq 0$ to in information about the degree of lity when $\theta=0$, as discussed in the section.

ermining Degree of Stability

ferring to equation 46, $f(\theta)$ is roughly as follows: Since

$$\frac{x(t)|}{} = a \cos\theta - b \sin\theta \tag{47}$$

for large r:

$$a \cos\theta - b \sin\theta \tag{48}$$

seen that if the system is stable then $=0$, $f(\theta)\approx a<0$. An angle $\theta=\theta_c\neq 0$ s such that $f(\theta_c)>0$. We now show to find this θ_c graphically from the

ppose that the CSH is in the LHP, own in Fig. 1. The critical angle θ_c is the direction of the ray which projection of a singularity at the . This angle θ_c measures the tion when first the system has an unded complex time response. To mine this angle first plot the CSH in -plane. Construct the perimeter of SH by connecting points of singu- with straight-line segments. Draw aight-line tangent to the perimeter own by b in Fig. 1) from the ori- Next construct a line or perpen- r to line $b0$. The inclination of line $0r$ is θ_c.

peat this procedure for the tangent in the lower-half plane $\omega<0$. The so obtained is the negative critical to be denoted by θ_c^1. The sum of these angles is the total critical angle θ_T. This leads to the definition

$$\eta = \frac{\theta_T}{\pi} \tag{49}$$

as the degree of stability.

This equation gives a first-order measure of the degree of stability. Since $\theta_T \leq \pi$, $\eta \leq 1$. When $\eta=1$ the system is non-oscillatory and exponentially damped. To compare two such systems a second-order measure of the degree of stability might be the magnitude of the displacement of CSH along the σ axis. This would be the magnitude of the distance of the nearest singularity to the origin lying on the real axis σ. A third measure would be the distribution of the projected singular points along the critical ray.

Conclusions

In this paper we have presented some aspects of a stability theory and a generalized stability criterion for the analysis of a certain class of nonlinear systems.

In nonlinear systems the stability is markedly dependent upon the boundary values (the initial conditions, for example). It is also dependent upon the magnitude of the driving function. In general, the range of amplitudes of the driving function and the range of initial or boundary conditions would have to be specified. A stability criterion for nonlinear systems must involve all these conditions in addition to the characteristics of the system. The proposed criterion based upon the concept of the convex singularity hull contains all these conditions automatically.

Because of the conformal properties of the CSH and the conjugate indicator diagram, these diagrams can be superimposed and relations between the frequency and time domains of a given nonlinear system obtained immediately.

In references 1, 2, 3 it is shown how a nonlinear differential equation such as equation 3 can be transformed to a recurrence equation. The latter can then be solved in open form because of its convolution and symmetry properties. In many instances the recurrence solution can be converted to a closed form by induction. Then the present method is immediately applicable. It appears from the Taylor-Cauchy inversion transform reported in reference 3 that it may be possible to ascertain the stability of a nonlinear system directly from the recursion formulas.

If the closed form of h_n in equations 1 and 2 is not known, it may be possible to obtain roughly the stability of the system by taking only the first few terms of the expansion of $x(t)$. In this case we must conceive a new function which majorizes $x(t)$ at least in the known truncated portion. However, caution must be taken to avoid serious errors if the unknown or discarded part is not majorized by this new function.

References

1. Recurrence Relations in the Solution of a Certain Class of Nonlinear Systems. **Alfred A. Wolf.** *AIEE Transactions*, vol. 78, pt. I (*Communication and Electronics*), 1959 (Paper no. 59-919).

2. A Mathematical Theory for the Analysis of a Class of Nonlinear Systems. **A. A. Wolf.** *Ph.D. Dissertation*, University of Pennsylvania, Philadelphia, Pa., June 1958.

3. Taylor-Cauchy Transforms for Solving Certain Class of Nonlinear Systems, **Y. H. Ku, A. A. Wolf.** *Convention Record*, Institute of Radio Engineers, New York, N. Y., 1959.

4. Entire Functions (book), **R. P. Boas.** Academic Press, New York, N. Y., 1954.

5. Theorie und Undwendung der Laplace Transformation (book), **G. Doetsch.** Dover Publishers, New York, N. Y., 1943.

Discussion

Tsang (University of Mississippi rsity, Miss.): The authors are to be atulated for dealing thoroughly with orny subject of stability of nonlinear ns. In fact the criterion is applicable function, not limited to those derived nonlinear systems.

concept of the CSH is introduced. e LHP is not a concave region, it that a simpler statement "all the arities must be in the LHP" should ficient for our purpose.

re will be considerable difficulty both iving the Laplace Transform from nlinear equations with their boundary ions, and in determining the singular from a Laplace Transform in inverse series of s. Equation 41 from equations 19, 20 with $x(0)=0$, $x'(0)=1$ is not typical for nonlinear systems. It would be very enlightening if the authors will apply the method in detail (first obtain a power series, then its Laplace Transform term by term and the singularities) to equations 19, 20 with the boundary conditions, say $x(0)=-1.9$, $x'(0)=0.2$.

Y. H. Ku and A. A. Wolf: The authors would like to thank Professor N. F. Tsang for his interesting comments on our paper, and take this opportunity to thank every one who discussed this paper with us at its presentation.

With regard to Professor Tsang's comment on the singularities, reasons shall be briefly given for the necessity of convexity of the singularity hull. The rate of growth of an entire function in different directions in the complex t-plane can be specified as indicated in reference 2 of the paper by the Phragmén-Lindelöf indicator function

$$fh(\theta) = \overline{\lim_{r\to\infty}} \frac{\log|x(re^{j\theta})|}{r^\rho} \tag{50}$$

where ρ denotes the order of the function and $t=re^{j\theta}$. It is known from the theory of entire functions that if $\rho \leq 1/2$ for $x(t)$ and if $x(t)$ is not a constant the minimum modulus is everywhere unbounded. That is to say that $x(t)$ denotes the response of an unstable system. In reference 2 of the paper it is shown that for certain class of physical systems ρ is unity or approximately so. Nonlinear physical systems discussed in this paper are governed by functions of exponential type and order one.

Putting $\rho=1$ in equation 50 gives the indicator function

$$fh(\theta) = \overline{\lim_{r\to\infty}} \frac{\log |x(re^{i\theta})|}{r} \tag{51}$$

for this class of physical systems.

The indicator function is everywhere bounded by the type of $x(t)$, namely τ. The type τ of a function $x(t)$ is given by

$$\tau = \lim_{r\to\infty} \frac{\log M(r)}{r} \tag{52}$$

according to reference 2 of the paper, where $M(r)$ denotes the maximum modulus of $x(t)$ and $r=|t|$.

Hence

$$M(r) = \max_{r\epsilon D_1} |x(t)| \tag{53}$$

where D_1 is the domain of convergence of $x(t)$ and r belongs to D_1. From equation 53 it is evident that

$$M(r) \geqslant |x(t)| \tag{54}$$

for all r belonging to D_1. Thus on substituting equation 54 into equation 52 and comparing the result with equation 51 one obtains the simple relation

$$fh(\theta) \leqslant \tau \tag{55}$$

Geometrically equation 55 signifies that $fh(\theta)$ is associated with a convex domain. Because of the conformal behavior of the complex t-plane and the complex s-plane, the distribution of singularities is also convex. Hence the concept of a CSH is important and useful.

This approach can be extended to those systems giving rise to nonconvex singularity hulls by suitably breaking the nonconvex sets into the sum of convex sets and then applying to each such convex set the theory presented in this paper.

In answering Professor Tsang's next comment, the example given in the paper shall be discussed with general boundary conditions:

$$x(0) = a = K_0;\ x'(0) = b = K_1 \tag{56}$$

where a and b are real constants. Equation 23 in the paper gives then

$$x(t) = \sum_{n=0}^{\infty} C_n \frac{t^{n+2}}{(n+1)(n+2)} + a + bt \tag{57}$$

The general recursion formula for C_n is given by

$$C_n = \frac{(-1)^n}{n!}[4^n - 2(3)^n] - \frac{2C_{n-1}}{n} - 2bC_n^{(0)} - \frac{C_{n-2}}{n(n-1)} - aC_n^{(0)} - bC_{n-1}^{(0)} - C_{n-4,0}^{(0)} - 2a\frac{C_{n-2}}{n(n-1)} - 2b\frac{C_{n-3}}{(n-1)(n-2)} - a^2C_n^{(0)} - 2abC_{n-1}^{(0)} - b^2C_{n-2}^{(0)} \tag{58}$$

where

$$C_m^{(0)} = \begin{cases} 1 & \text{when } m=0 \\ 0 & \text{otherwise} \end{cases} \tag{59}$$

$$C_{n-4,0}^{(2)} = \sum_{k=0}^{n-4} \frac{C_k C_{n-k-4}}{(k+1)(k+2)(n-k-3)(n-k-2)} \tag{60}$$

When $a=0$ and $b=1$, one obtains the recurrence relation (see equation 60) from which equation 28 in the paper is obtained.

$$C_n = \frac{(-1)^n}{n!}[4^n - 2(3)^n] - \frac{2C_{n-1}}{n} - 2C_n^{(0)} - \frac{C_{n-2}}{n(n-1)} - C_{n-1}^{(0)} - C_{n-4,0}^{(2)} - \frac{2C_{n-3}}{(n-1)(n-2)} - C_{n-2}^{(0)} \tag{61}$$

It is always possible to solve equations 5[9] and 60 recursively; i.e., for $n=0, 1, 2, \ldots$ Suppose that this is done, then the nth term shall either be recognized or no[t]. In the former case one proceeds as in th[e] paper. In the latter case the form [of] $C_0, C_1, C_2, \ldots$ might be examined to se[e] if it takes a particular form. One suc[h] form might be

$$C_n = \frac{A\alpha^n}{n!} \tag{62}$$

where A and α are constants, as suggeste[d] by the nth term given in the paper. [If] this should prove to be unfruitful, instea[d] a majorizing or asymptotic solution f[or] C_n might be obtained.

In conclusion, it is seen that it is [a] straightforward matter to obtain recurren[ce] relations which can be solved. It is reco[g]nized that some difficulties may arise [in] getting the closed form of the nth term. Further efforts are being made in tha[t] direction. It may be remarked that th[e] method of stability analysis can be revers[ed] so that one can synthesize a nonline[ar] feedback control system with a prescrib[ed] frequency response. In this case the n[th] term is known as soon as the CSH i[s] specified. It is even possible to build in[to] the system a given degree of stability, [a] given settling time, and other releva[nt] characteristics.

Notes on Bridged-T Complex Conjugate Compensation and 4-Terminal Network Loading

PRAPAT CHANDAKET
NONMEMBER AIEE

ALLEN B. ROSENSTEIN
MEMBER AIEE

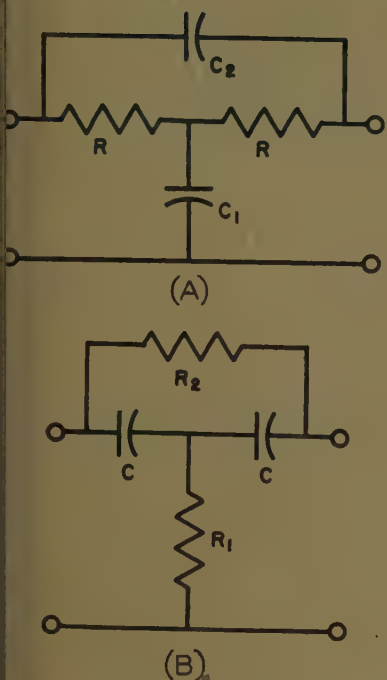
MANY servo systems possess open-loop transfer functions which contain one or more pairs of complex conjugate poles. Often a pair of these complex conjugate poles will lie close to the imaginary axis and will cause the uncompensated system to possess a relatively poor damping ratio or damping exponent. Usually, satisfactory compensation of such systems can be achieved through the use of tachometer feedback or the conventional compensators possessing real axis singularities. However, there are occasions when complex conjugate singularity compensation would possess distinct advantages. This is particularly true if the compensating complex singularities can be introduced by a simple R-C (resistance-capacitance) passive network with zero d-c attenuation.

Evans[1] has shown how the simple bridged-T network can be used with a-c servos to produce a phase lead. In this article, the application of the elementary bridged-T network is developed for d-c system compensation. The locus of roots of the transfer function of t[he] bridged-T network is presented in dimen[-]sionless form. It is found that t[he] transfer function of the bridged-[T] possesses two zeros, one or the oth[er] of which may be adjusted to fall upon a[ny] desired point in the left half-plane i[n]cluding any complex conjugate pa[ir]. The bridged-T also produces a pair [of] poles which fall upon the negative r[eal] axis on either side of the real componen[t] of the pair of zeros.

A set of curves has been develop[ed] which allows very rapid synthesis of t[he] unloaded bridged-T network with [the] compensating complex conjugate zer[os] falling at any desired left half pla[ne]

Paper 59-220, recommended by the AIEE Feedb[ack] Control Systems Committee and approved by [the] AIEE Technical Operations Department [for] presentation at the AIEE Winter General Meeti[ng,] New York, N. Y., February 1–6, 1959. Manuscr[ipt] submitted October 2, 1958; made available [for] printing December 16, 1958.

LIEUT.-COMDR. PRAPAT CHANDAKET is with [the] Royal Thai Navy, and ALLEN B. ROSENSTEI[N is] with the University of California, Los Ange[les,] Calif.



. 1. **Common types of bridged-T configuration**

nts. Both types of simple bridged-T work can be synthesized from the ne graphs.

To investigate the effects of loading on the bridged-T network, a general thod has been developed to observe locus of the transfer function poles d zeros of any four terminal passive work with varying load termination. e effects of three different types of dged-T loading are investigated. It is nd that these loads do not affect the ation of the zeros but can be employed change materially the location of the mpensating poles upon the negative l axis.

The application of the loaded bridged-T work to a position servomechanism, h as a fin control servomechanism ose open-loop transfer function along h the airplane dynamics function, tains complex poles in the left halfne, is found to offer material improvents in the over-all system performces.

alysis of the Bridged-T Network

ANSFER FUNCTION

Bridged-T *R-C* networks fall in two es of configurations; See Fig. 1. The work in (A) will be referred to as dged-T type *Ia*, and the other as type Bridged-T type *Ia* has been anaed by Evans[1] and others. Evans used t locus techniques to determine its e-zero configuration and found it to be characterized by two poles and two zeros. These poles are always real and negative, but the zeros can be made negative real or complex, with negative real parts depending on the ratio of capacitances.

The root locus technique will be used to show the movement of the poles and zeros of the transfer functions of these networks. The bridged-T type *Ib* is selected for illustration. With zero input source impedance and infinite output load impedance, the voltage transfer function of the network type *Ib* is found to be

$$\frac{E_{out}}{E_{in}}=\frac{1+2R_1Cs+R_1R_2C^2s^2}{1+(2R_1+R_2)Cs+R_1R_2C^2s^2} \qquad (1)$$

LOCUS OF POLES AND ZEROS

The equation for the zeros of equation 1 is given by

$$1+2R_1Cs+R_1R_2C^2s^2=0 \qquad (2)$$

which can be arranged as

$$\frac{1+2R_1Cs}{R_1R_2C^2s^2}=-1 \qquad (3)$$

The pole equation can also be arranged as

$$\frac{1+2R_1Cs}{R_2Cs(1+R_1Cs)}=-1 \qquad (4)$$

If $1/R_2$ is assumed to be the loop gain and allowed to vary from zero to infinity, the root locus plot of equation 3 will give the zero configuration for the network, while the root locus of equation 4 will yield the pole characteristics. These are shown in Fig. 2.

From Fig. 2, it may be seen that the zeros of the transfer function can be made complex or real in the left half-plane while the poles are always negative real when R_2 is varied.

SIMILARITIES

The transfer function of the network type *Ia* with zero input impedance and infinite output load impedance is found to be

$$\frac{E_{out}}{E_{in}}=\frac{1+2RC_2s+R^2C_1C_2s^2}{1+R(C_1+2C_2)s+R^2C_1C_2s^2} \qquad (5)$$

Comparing the transfer functions of both networks, equations 1 and 5, it is obvious that they have similar characteristics. If R, C_1, C_2 (type *Ia*), equation 4, are replaced by C, R_2 and R_1 respectively, equation 4 becomes the transfer function for type *Ib* as shown by equation 1. A single graph giving the pole-zero laws for both types of network in normalized coordinates would be a convenient design tool. Such a graph is developed in the next section to allow the rapid location of a pair of complex conjugate zeros anywhere in the left half-plane.

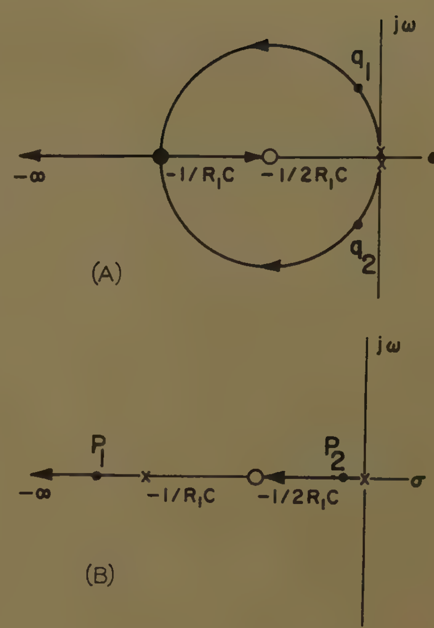


Fig. 2. Root loci of zeros and of poles of transfer function versus loop gain $1/R_2$

A—Of zeros
B—Of poles

Normalized Characteristics

For design convenience, a graphical method is developed in this section. It should be made clear at the beginning that all derivations are given with no loading. It is desired to design the bridged-T networks with specified conjugate complex zeros. The poles are dedependent on the zero locations. However, loading influences the pole-zero location of the transfer function. This effect will be discussed in the next section, where it will be used as part of the synthesis procedure.

NORMAL DESIGN CHART DEVELOPMENT

As shown, type *Ia* and *Ib* networks are equivalent to each other. The same transfer function can be obtained just by switching their resistances and capacitances. Therefore, for the development of the design charts, only network *Ib* will be analyzed.

As shown in Fig. 2(A), the locus of complex zeros is along the circle centered at $-(1/2R_1C)$. If equation 2 is rewritten in terms of its damping ratio and undamped natural frequency (ω_n) as follows:

$$1+2R_1Cs+R_1R_2C^2s^2=\frac{1}{R_1R_2C^2}+\frac{2}{R_2C}s+s^2=\omega_n^2+2\zeta\omega_ns+s^2=0 \qquad (6A)$$

then

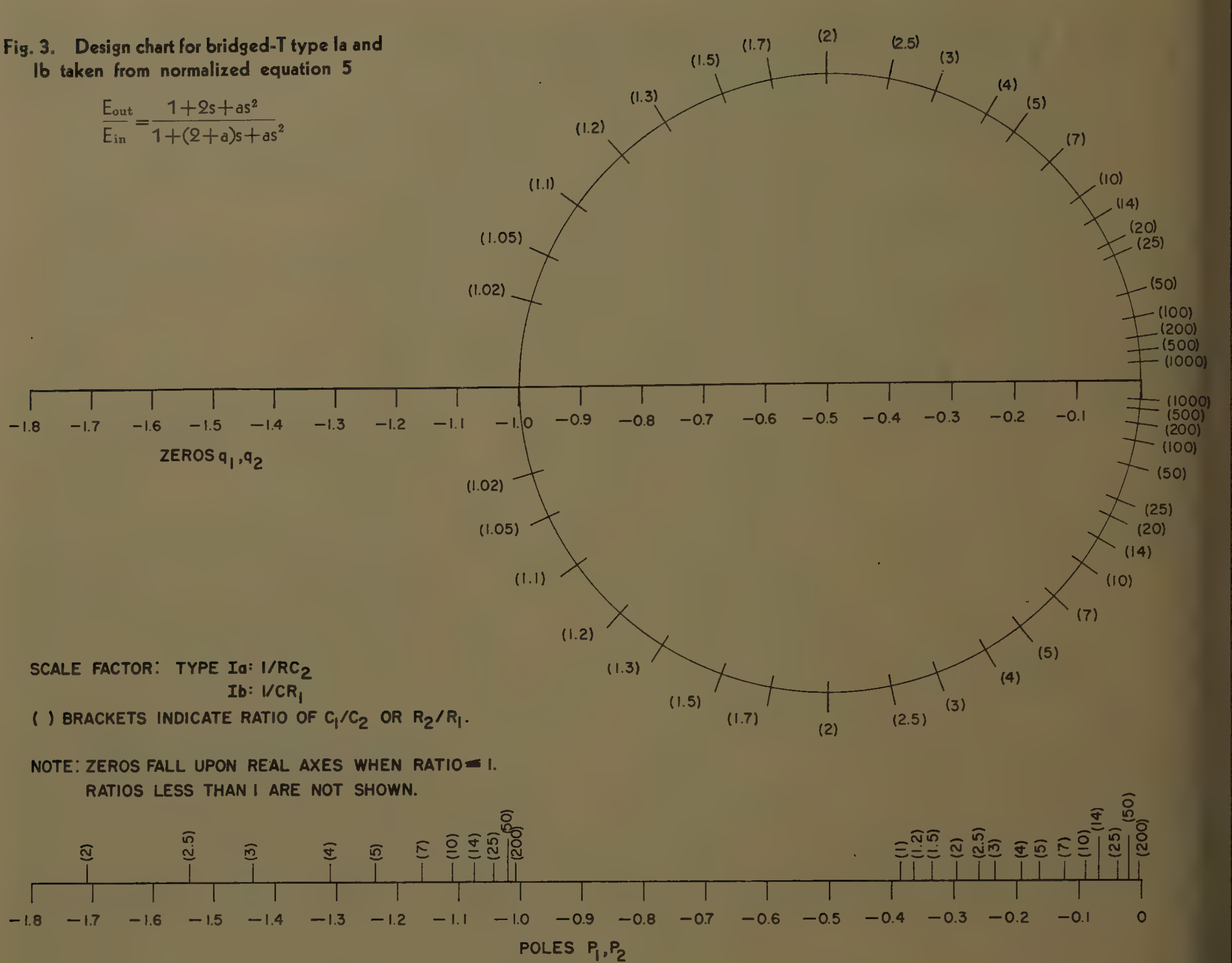


Fig. 3. Design chart for bridged-T type Ia and Ib taken from normalized equation 5

undamped frequency $=\omega_n=\frac{1}{C\sqrt{R_1R_2}}$

damping ratio $=\zeta=\sqrt{R_1/R_2}$ (6B)

and

$\omega_n/\zeta=1/R_1C$ (6C)

The damping ratio ζ is seen to depend upon the ratio, R_2/R_1. To normalize the locus of zeros, Fig. 2(A), a scale change can be introduced to cause the circle diameter to be equal to unity or, in other words, to cause $1/R_1C=1$. With $R_1C=1$, equation 2 is normalized and can be rewritten

$$\frac{1}{R_1R_2C^2}+\frac{2}{R_2C}s+s^2=\frac{R_1}{R_2}+2\frac{R_1}{R_2}s+s^2$$
$$=\omega_{nz}^2+2\zeta_{nz}\omega_{nz}s+s^2=0 \quad (7)$$

from which it may be seen that

normalized natural frequency

$$=\omega_{nz}=\sqrt{\frac{R_1}{R_2}}=\frac{1}{\sqrt{a}} \quad (8A)$$

normalized damping ratio

$$=\zeta_{nz}=\sqrt{R_1/R_2}=\frac{1}{\sqrt{a}} \quad (8B)$$

if it is assumed that $a=R_2/R_1$.

But from equations 6(B), 8(A) and 8(B), it is noticed that the original damping ratio, the normalized damping ratio, and the normalized natural frequency of oscillation are all equal.

$$\zeta=\zeta_{nz}=\omega_{nz}=\sqrt{R_1/R_2}=\frac{1}{\sqrt{a}} \quad (9)$$

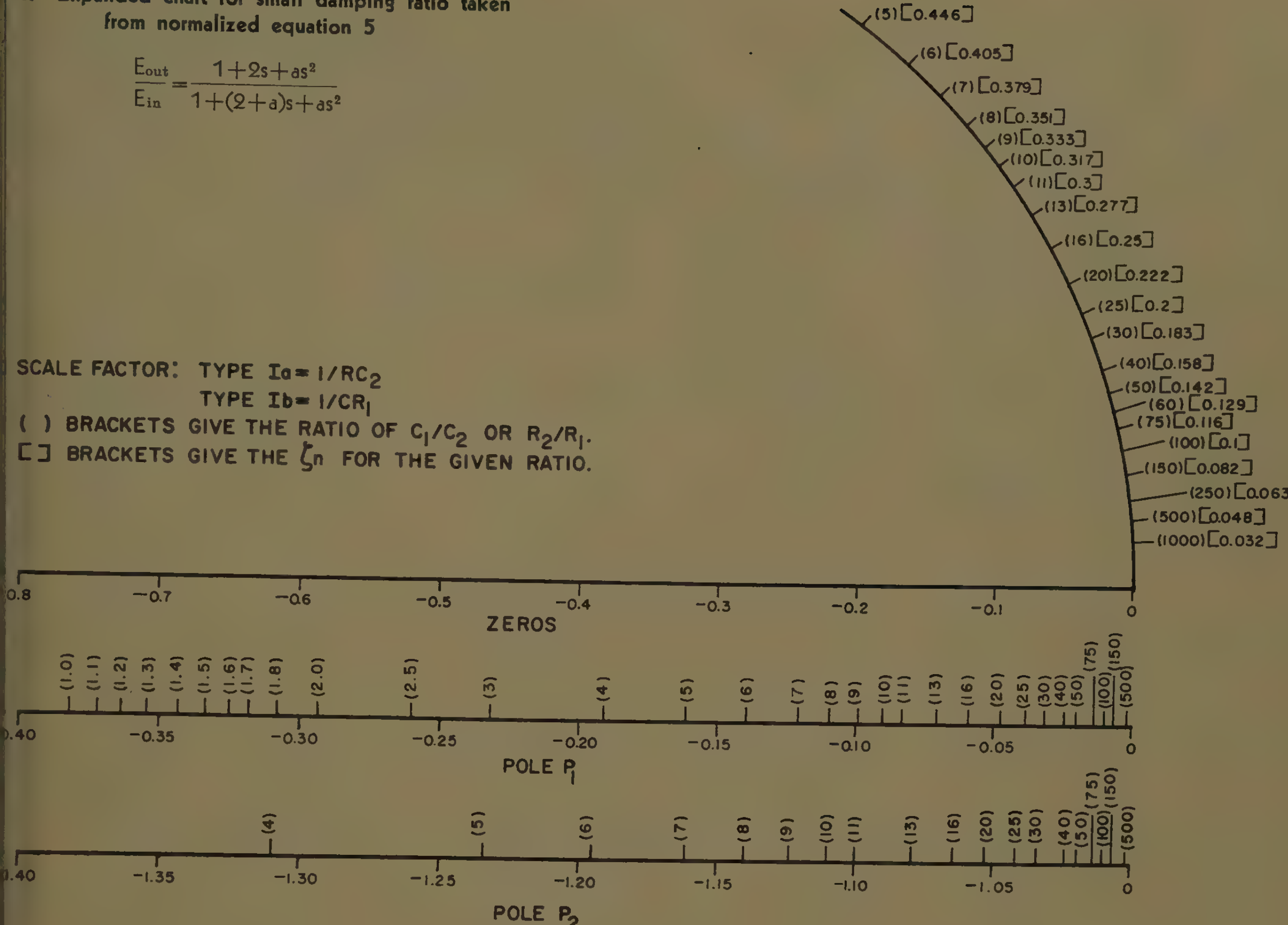
For complex conjugate zeros, it is necessary for R_2 to be greater than R_1. Normalizing equation 2 into equation 7 by setting $1/R_1C=1$, the locus of zeros can be plotted, Fig. 3, as the ratio of R_2/R_1 is varied from unity to infinity. Similarly, equation 4 can also be normalized with the use of the scale factor R_1C and the locus of poles plotted in Fig. 3 as the ratio R_2/R_1 is varied from unity to infinity. Note that the plots could have been presented in terms of R_1/R_2, but it has been considered desirable to use large numbers rather than small decimals.

The ability of the bridged-T network to synthesize any left half-plane complex zero can be realized from Fig. 3. Here it can be seen that complex zeros can be readily placed with any desired damping ratio between zero and unity. Further since the damping ratio is invariant with t scale change, the designer has only to change scale by selecting the proper value of $1/R_1C$ to adjust the complex zeros to any desired values on the required ζ line.

The chart of Fig. 3 can be used for networks of both types *Ia* and *Ib*. The numbers in the bracket indicate the ratio C_1/C_2 for network *Ia* or R_2/R_1 for network *Ib*. Direct readings of poles and zeros can be obtained only under the condition that $1/RC_2$ (network *Ia*) or $1/CR_1$ (network *Ib*) equal unity. The exact location of the poles and zeros can be obtained by multiplying the chart reading by $1/RC_2$ or $1/CR_1$. These factors are therefore the mentioned scale factors for the determination of the actual pole-zero locations. On the other hand, the desired locations of the zeros might be given and the problem would then be one of determining the required network. In this case, the given value of ζ would produce R_2/R_1 or C_1/C_2 from the chart. The ratio between the actual co-ordinate of the zeros and the normalized co-ordinates

4. Expanded chart for small damping ratio taken from normalized equation 5

$$\frac{E_{out}}{E_{in}}=\frac{1+2s+as^2}{1+(2+a)s+as^2}$$



ld produce the scale factor which s $1/RC_2$ or $1/R_1C$.

or compensation in many practical lications, one must place the com- zeros close to the imaginary axis. obtain more accuracy, an expanded t of the region in the vicinity of the ginary axis has been provided in Fig.

SIGN CHART APPLICATION

wo examples of the application of the ged-T charts are given:

xample 1: To determine of the volt- transfer function of the network shown Fig. 5(A). This is a type *Ia* network $1/RC_2=10$, and $C_1/C_2=10$. $1/RC_2$ omes the scale factor. From the chart Fig. 4:

s: $-0.1\pm j0.3$
s: -0.09, -1.11

Multiply singularities by $1/RC_2$ to ob- the correct transfer function.

$$=\frac{(s+1+j3)(s+1-j3)}{(s+0.9)(s+11.1)}$$

xample 2: To design a bridged-T *Ia* network to give a pair of complex s at $\zeta=0.2$ and $\omega_n=2$, and then find its voltage transfer function.

From equation 6(B): If $\zeta=0.2$, then R_2/R_1 or $C_1/C_2=25$

From the expanded chart, Fig. 4, for $C_1/C_2=25$

Zeros: $-0.04\pm j0.195$
Poles: -0.0384, -1.042

From equation 6(C),

$$1/RC_2=\omega_n/\zeta=2/0.2=10$$

Actual pole-zero location:
Zeros: $-0.4\pm j1.95$
Poles: -0.384, -10.42

Transfer function

$$=\frac{(s+0.4+j1.95)(s+0.4-j1.95)}{(s+0.384)(s+10.42)}$$

Network type *Ia* can be obtained by selecting its components under the condition that $C_1/C_2=25$ and $1/RC_2=10$. If one selects $R=1$ megohm, then $C_2=0.1$ microfarad and $C_1=2.5$ microfarads. This network is shown in Fig. 5(B).

SUMMARY OF PROPERTIES OF UNLOADED BRIDGED-T

From the previous analysis, the properties of both types of networks without loading may be summarized as follows. Refer to Fig. 1(A) and 1(B):

1. Product of poles = product of zeros (see equations 1 and 5)

2. Poles are always on negative real axis.

3. Zeros can be made negative real or complex according to the ratio of C_1/C_2 (*Ia*) or R_2/R_1 (*Ib*).

Network *Ia:*

$C_1/C_2>1$, complex
$C_1/C_2=1$, double, negative real
$C_1/C_2<1$, different, negative real

Network *Ib:*

$R_2/R_1>1$, complex
$R_2/R_1=1$, double, negative real
$R_2/R_1<1$, different, negative real

4. One pole is always closer to the origin than the real part of the zeros.

5. On the complex plane, the locus of zeros is a circle.

6. No d-c attenuation.

7. For very large C_1/C_2 or R_2/R_1, one pole and two zeros approach the origin; another pole approaches $-1/RC_2$ for type *Ia* and $-1/CR_1$ for type *Ib*.

APPLICATION

The property *g* is sometimes very useful in stabilization of a certain type of d-c servos, such as those positioning types with an inner loop pole at the origin and two more poles on the negative real axis. Suppose the open-loop transfer function of this system is $K/[s(1+\tau_1 s)(1+\tau_2 s)]$ In this case, the pole at the origin will associate with the two complex zeros and the one negative real pole which are all close to the origin; see Fig. 6.

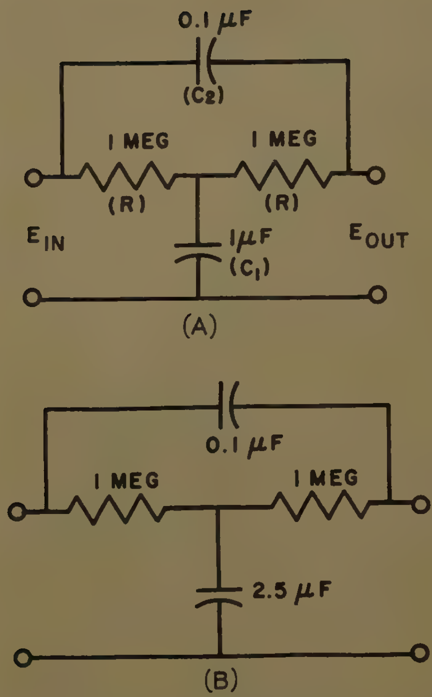


Fig. 5. Network configurations. A—For example 1. B—For example 2

If RC_2 (for type *Ia*) or CR_1 (for type *Ib*) is made very small, then the other pole of the bridged-T transfer functions will be very far from the origin. It will therefore not be significant in controlling the system roots. Those singularities which group together very close to the origin will have little effect on the two remaining poles $-1/\tau_1$ and $-1/\tau_2$. Consequently the root locus with a break between these two remaining poles will be practically vertical and one can increase the gain to a very high level while maintaining a large damping exponent for the predominant roots. The main limitation for this application of the bridged-T is the roots near the origin. However, if the gain can be made very high with the system stability still satisfied, then the residues due to the roots close to the origin will be very small, and will contribute very little to the closed-loop transient response. This example, however, is somewhat oversimplified and must be used with caution. The presence of the zeros in the vicinity of the origin will introduce a low-frequency notch in the open loop transfer function.

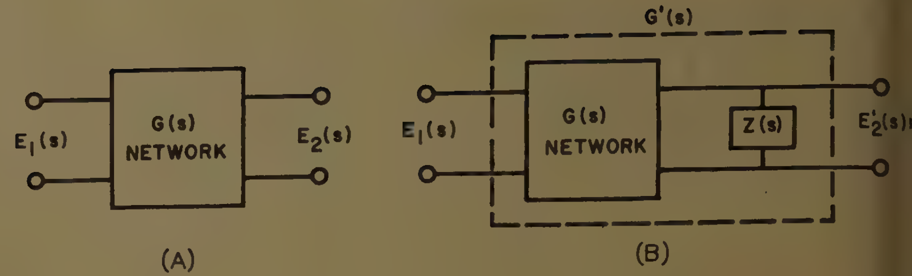


Fig. 7. Transfer function G(s) of four terminal passive networks

Effects of Loading Impedance

It is known that, if a 4-terminal network is terminated by any load impedance, the pole-zero configuration of its original unloaded transfer function will be affected and one will obtain a new configuration. It was desirable, therefore, to develop a technique to observe the change due to this loading effect. Fortunately, this technique can be used as part of the synthesis procedure in the bridged-T design. Application of this concept will be illustrated in the stabilization of a control loop containing airplane dynamics in the following section.

Generally, this method can be applied to any 4-terminal network to observe the locus of the pole-zero change of the original transfer function when a variable load impedance is connected across its output terminals. It can be used, as well, for adjustment of the configuration.

Theory

The method employs Thevenin's theorem in conjunction with root locus. The pole-zero change of the network transfer function due to loading can be obtained by the following steps:

1. Having determined the original unloaded system transfer function, use Thevenin's theorem to find a new transfer function with load impedance $Z(s)$ being placed across the output terminal.

2. Convert this function into an equivalent feedback control system.

3. Apply root locus technique to the equivalent control system in step 2. The roots of the locus plot will now show the changing poles and zeros under the variable load conditions.

The three steps will be analyzed as follows: Suppose

$G(s)$ is the original unloaded voltage transfer function.

$Z(s)$ is the load impedance function.

$G'(s)$ is the new transfer function including load impedance $Z(s)$.

From Fig. 7(A)

$$\frac{E_2(s)}{E_1(s)}=G(s) \tag{10}$$

From Fig. 7(B)

$$\frac{E_2'(s)}{E_1(s)}=G'(s) \tag{11}$$

By Thevenin's theorem, Fig. 7(B) can be transformed to Fig. 8. The new transfer function is

$$G'(s)=\frac{E_2'(s)}{E_1(s)}=\frac{G(s)Z(s)}{Z'(s)+Z(s)} \tag{12}$$

where $Z'(s)$ = impedance looking back into the output of network $G(s)$ with the input short-circuited.

Suppose $G(s)$, $Z(s)$, $Z'(s)$ are rewritten as the quotient of polynomials as follows:

$$G(s)=\frac{G_a(s)}{G_b(s)}$$

$$Z(s)=\frac{Z_a(s)}{Z_b(s)}$$

$$Z'(s)=\frac{Z_a'(s)}{Z_b'(s)} \tag{13}$$

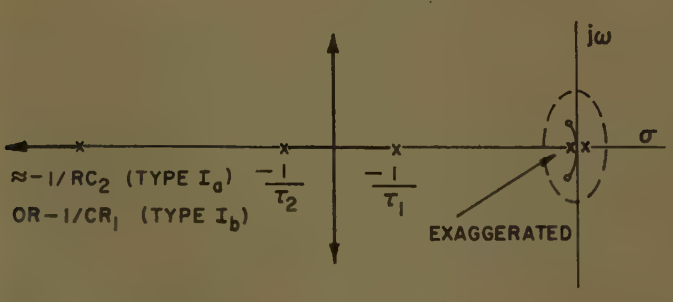


Fig. 6. Root locus plot of type I d-c servo with bridged-T network at extreme conditions as its compensating device

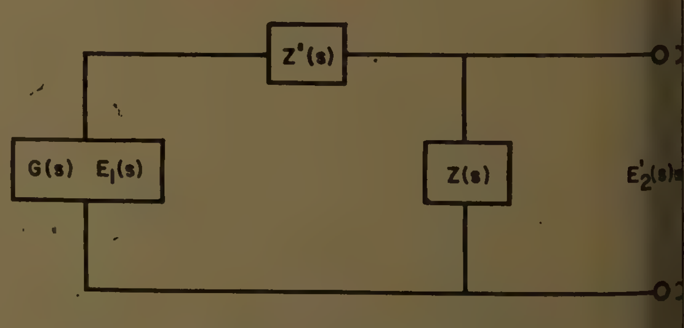


Fig. 8. Transformed circuit by Thevenin's theorem

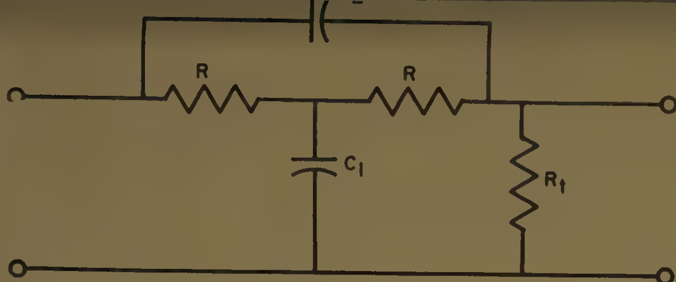


Fig. 9. Bridged-T type Ia with load resistance R_t

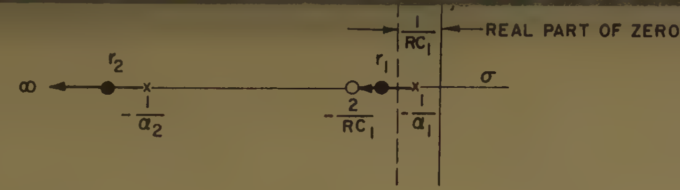


Fig. 10. Root loci for

$$\frac{2R}{R_t}\frac{(1+RC_1s/2)}{(1+\alpha_1s)(1+\alpha_2s)}$$

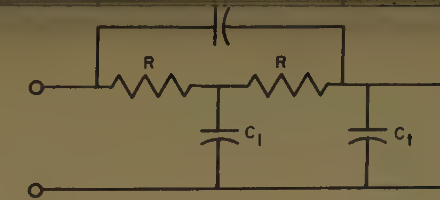


Fig. 11. Bridged-T type Ia with capacitive load

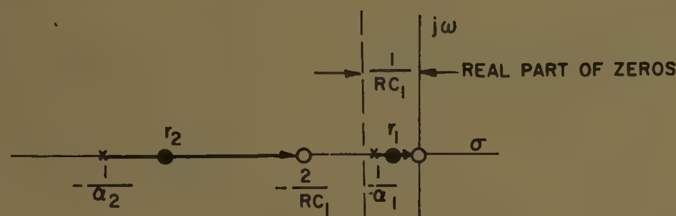


Fig. 12. Root loci of

$$2RC_t\frac{s(1+RC_1s/2)}{(1+\alpha_1s)(1+\alpha_2s)}$$

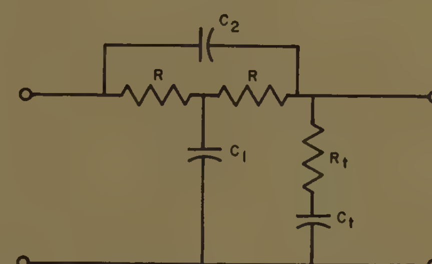


Fig. 13. Bridged-T type Ia with $Z(s)=R_t+1/C_ts$

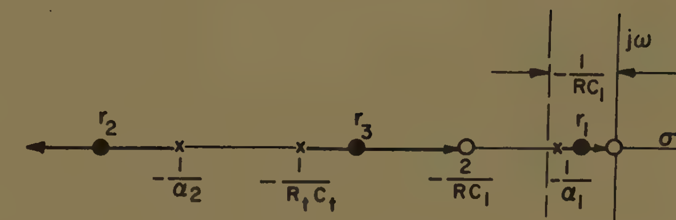


Fig. 14. Root loci for

$$2RC_t\frac{s(1+RC_1s/2)}{(1+\alpha_1s)(1+\alpha_2s)(1+R_tC_ts)}$$

as a function of C_t with R_tC_t constant

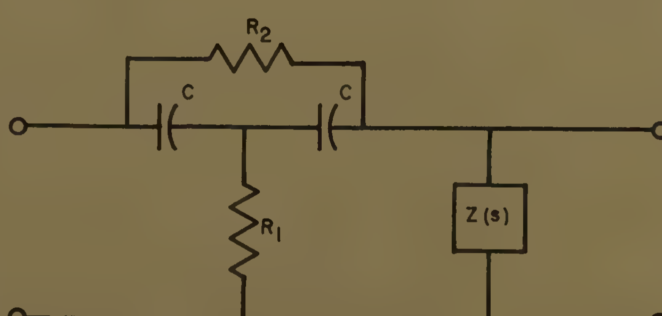


Fig. 15. Bridged-T type Ib with load impedance Z(s)

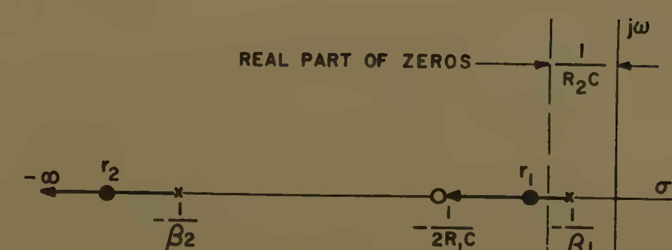


Fig. 16. Root loci for

$$\frac{R_2}{R_t}\frac{(1+2R_1Cs)}{(1+\beta_1s)(1+\beta_2s)}$$

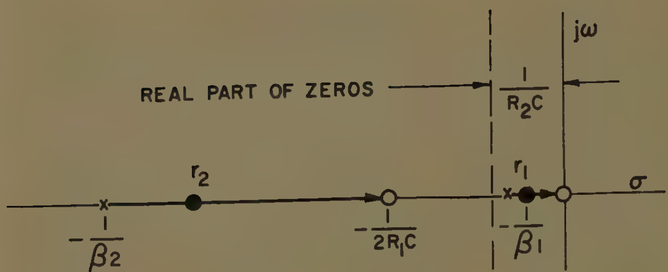


Fig. 17. Root loci for

$$C_tR_t\frac{s(1+2R_1Cs)}{(1+\beta_1s)(1+\beta_2s)}$$

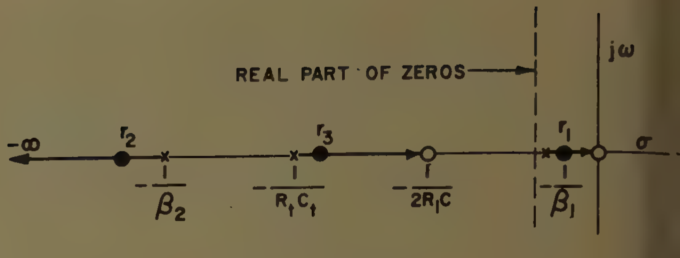


Fig. 18. Root loci for

$$C_tR_2\frac{s(1+2R_1Cs)}{(1+\beta_1s)(1+\beta_2s)(1+R_tC_ts)}$$

Note that $Z_b'(s)=G_b(s)$ by the property of all four terminal passive networks. Then, from equation 12

$$G'(s)=\frac{G(s)}{1+Z'(s)/Z(s)} \tag{14}$$

Then the poles and zeros of the new function $G'(s)$ can be easily obtained as follows:

1. Zeros of $G'(s)$=combined zeros of $G(s)$ and $Z(s)$.
2. Poles of $G'(s)$ are the roots of $[Z'(s)]/[Z(s)]=-1$, which can be found by the root locus technique.
3. D-c attenuation$=G'(0)$ which can be obtained from equation 14.

Finally, $G'(s)$ can be written as

$$G'(s)=G'(0)\frac{(1-s/a_1)(1-s/a_2)\ldots(1-s/a_n)}{(1-s/r_1)(1-s/r_2)\ldots(1-s/r_n)} \tag{15}$$

where

$a_1, a_2, \ldots a_n$ are zeros of $G(s)$ and $Z(s)$

$r_1, r_2, \ldots r_n$ are roots of $\frac{[Z'(s)]}{[Z(s)]}=-1$

Loading Effect on Type Ia

To illustrate the application of this method, three types of load impedance will be applied first to bridged-T type *Ia*. These are pure resistance, pure capacitance, and a combination of resistance and capacitance.

Case Ia: $Z(s)$ is pure resistance; see Fig. 9. Suppose

$$G(s)=\frac{1+as+bs^2}{(1+\alpha_1s)(1+\alpha_2s)}$$

$$Z(s)=R_t$$

$Z'(s)$ is found to be

$$Z'(s)=\frac{2R(1+RC_1s/2)}{1+R(C_1+2C_2)s+R^2C_1C_2s^2} \tag{16}$$

or

$$Z'(s)=\frac{2R(1+RC_1s/2)}{(1+\alpha_1s)(1+\alpha_2s)}$$

$$\frac{Z'(s)}{Z(s)}=\frac{2R}{R_t}\frac{(1+RC_1s/2)}{(1+\alpha_1s)(1+\alpha_2s)}=-1 \tag{17}$$

A sketch of the root loci for poles of $G'(s)$ is shown in Fig. 10 by equating $Z'(s)/Z(s)=-1$.

Finally,

$$G'(s)=\frac{R_t}{2R+R_t}\frac{1+as+bs^2}{(1-s/r_1)(1-s/r_2)} \tag{18}$$

Analysis of Bridged-T type *Ia* with load resistance R_t shows the following properties:

1. D-c attenuation$=R_t/(2R+R_t)$.
2. Zeros are the same as original.
3. Both poles of $G'(s)$ are real and these move to the left if R_t is decreased.
4. Limit of poles movement for the one near the origin is $-2/RC_1$ or twice the real part of zeros. The other pole will move towards infinity along the negative real axis.

It is of interest to note that, with the introduction of d-c attenuation, the poles of the original bridged-T transfer function can be pulled farther into the left half-plane with the pole at $-1/\alpha_1$, and are even capable of moving to the left of the real part of the zeros.

Case IIa: $Z(s)$ is pure capacitance; see Fig. 11. Assume that $G(s)$ and also $Z'(s)$ are the same as in case *Ia* but that $Z(s)=1/C_ts$. Then

$$\frac{Z'(s)}{Z(s)}=2RC_t\frac{s(1+RC_1s/2)}{(1+\alpha_1s)(1+\alpha_2s)}=-1 \tag{19}$$

The root loci of $Z'(s)/Z(s)=-1$ gives the locations of poles in $G'(s)$. The locations of the new poles of $G'(s)$ depend on C_t if $G(s)$ is fixed. The pole movement versus the loop gain C_t is shown in Fig. 12.

$$G'(0)=1$$

Then

$$G'(s)=\frac{1+as+bs^2}{(1-s/r_1)(1-s/r_2)}, \tag{20}$$

The analysis of bridged-T type *Ia* with pure capacitive load C_t shows the following properties:

1. It contains no d-c attenuation.
2. Zeros are the same as the original.
3. Both poles of $G'(s)$ are real and move to the right if C_t is increased.
4. When C_t approaches infinity, one pole will approach the origin and the other pole approaches $-2/RC_1$, which is twice the real part of zero.

Case IIIa: $Z(s)$ is a series combination of resistance and capacitance; see Fig. 13. $G(s)$ and $Z'(s)$ are taken the same as the previous two cases.

$$Z(s)=\frac{1+R_tC_ts}{C_ts}$$

Zeros of $G'(s)=(1+as+bs^2)(1+R_tC_tS)$

Poles of $G'(s)$=roots of equation

$$2RC_t\frac{s(1+RC_1s/2)}{(1+\alpha_1s)(1+\alpha_2s)(1+R_tC_ts)}=-1 \tag{21}$$

Root loci for expression 19 are shown in Fig. 14.

If R_t and C_t are arbitrarily selected, one zero in $G'(s)$ will be fixed at $-1/R_tC_t$; also, the poles of $G'(s)$ are fixed. Obviously, this network has no d-c attenuation. Suppose the roots of the system are fixed at r_1, r_2, and r_3; then

$$G'(s)=\frac{(1+as+bs^2)(1+R_tC_ts)}{(1-s/r_1)(1-s/r_2)(1-s/r_3)} \tag{22}$$

To make this analysis applicable, it should be emphasized that one negative real zero may be added at any point depending on the selection of R_t and C_t. Poles can be adjusted by varying C_t with the condition that R_tC_t be kept unchanged. Under this condition C_t (see equation 21) can be considered as the open-loop gain of the system.

Analysis of $G'(s)$ of bridged-T type *Ia* with series resistance R_t and capacitance C_t as load-impedance shows the following properties:

1. There is no d-c attenuation.
2. Contains three poles and three zeros (two original and one negative real zero at $-1/R_tC_t$).
3. For a specified position of new zeros poles can be adjusted by varying C_t with R_tC_t maintained constant.

Loading Effects on Type Ib

Analysis of bridged-T type *Ib* with load impedance can be obtained by the same technique used for type *Ia*. In this section, only the results and useful expressions will be given. Since the use of these networks will usually concern complex zeros, the zeros of $G(s)$ discussed

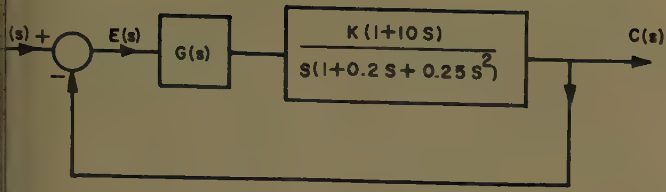


Fig. 19. Control system for example 1

this section are assumed to be conjugate complex; see Fig. 15.

et

$$s) = \frac{1+a's+b's^2}{(1+\beta_1 s)(1+\beta_2 s)} \quad (23)$$

All β_1, β_2, a', and b' are fixed and will used for all cases in later discussion in is section.

Note that the real part of the zeros $G(s)=1/R_2C$.

$$(s) = \frac{R_2(1+2R_1Cs)}{(1+\beta_1 s)(1+\beta_2 s)} \quad (24)$$

Case Ib: $Z(s)$ is pure resistance (R_t). eros of $G'(s)$ = zeros of $G(s)$, and poles $G'(s)$ = roots of

$$\frac{(1+2R_1Cs)}{(1+\beta_1 s)(1+\beta_2 s)} = -1 \quad (25)$$

See Fig. 16 for root loci of equation 25. R_t is fixed and the roots are at r_1 and r_2, en

$$(s) = \frac{R_t}{R_2+R_t}\,\frac{(1+a's+b's^2)}{(1-s/r_1)(1-s/r_2)} \quad (26)$$

Case IIb: $Z(s)$ is pure capacitance C_t. eros of $G'(s)$ = zeros of $G(s)$, and poles $G'(s)$ = roots of

$$R_2\,\frac{s(1+2R_1Cs)}{(1+\beta_1 s)(1+\beta_2 s)} = -1 \quad (27)$$

See Fig. 17 for root loci of equation 27. C_t is fixed, the roots are at r_1 and r_2, en

$$(s) = \frac{(1+a's+b's^2)}{(1-s/r_1)(1-s/r_2)} \quad (28)$$

Case IIIb: $Z(s)$ is the series combination of resistance (R_t) and capacitance $_t)$. Zeros of $G'(s)=(1+as+bs^2)(1+C_ts)$, and poles of $G'(s)$ = roots of

$$R_2\,\frac{s(1+2R_1Cs)}{(1+\beta_1 s)(1+\beta_2 s)(1+R_tC_ts)} = -1 \quad (29)$$

See Fig. 18 for root loci of equation 29. If R_t and C_t fix the roots at r_1, r_2, and r_3, then

$$G'(s) = \frac{(1+a's+b's^2)(1+R_tC_ts)}{(1-s/r_1)(1-s/r_2)(1-s/r_3)} \quad (30)$$

The properties of loading effect in these three cases for network type *Ib* are generally similar to that of type *Ia*. The differences are only the d-c attenuation terms and the limits which the roots of $[Z'(s)]/(Zs)] = -1$ will approach when the loop gain is increased towards infinity. Even in case *IIIb*, after another zero has been added, the poles must be adjusted (as in case *IIIa*) by varying C_t with R_tC_t maintained constant.

Use of Loading Effects

The behavior of the voltage transfer functions of both types of bridged-T networks under the three loading conditions have been discussed in the last two sections. One should note that, under the same loading condition, the two types of network behave in a similar manner. The only slight difference is the position of one zero in the equivalent open-loop transfer function $Z'(s)/Z(s)$ for the determination of poles in $G'(s)$. This zero will give a different termination of one root.

For example, in case *Ia*, this zero is at $-2/RC_1$ (see equation 17 and Fig. 10) and becomes the terminating zero for the root locus from the pole $-1/\alpha_1$. This is twice the real part of complex zeros of $G(s)$. In case *Ib*, (see equation 23 and Fig. 15) this zero is at the center of the circle passing through complex zeros of $G(s)$ and the origin. Consequently, even though $G(s)$ for networks *Ia* and *Ib* are the same, one will not obtain the same pole locations in $G'(s)$ for equivalent load resistance R_t.

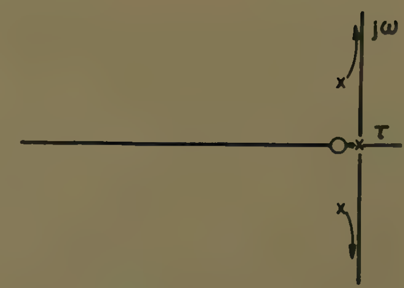


Fig. 20. Sketch of root loci for $\dfrac{K(1+10s)}{s(1+0.2s+0.25s^2)}$

It has been shown that the damping ratio ζ of the zeros in $G(s)$ is controlled by R_2/R_1 for type *Ib* and C_1/C_2 for type *Ia*. If the $G(s)$ in both types are the same, then with very small ζ the limit to which the pole $-1/\beta_1$ (type *Ib*) can move away to catch a zero at $-1/(2R_1C)$ is greater than that to which the pole $-1/\alpha_1$ (type *Ia*) can move to catch a zero at $-2/RC_1$. Thus, if ζ is large, then type *Ia* would be preferred.

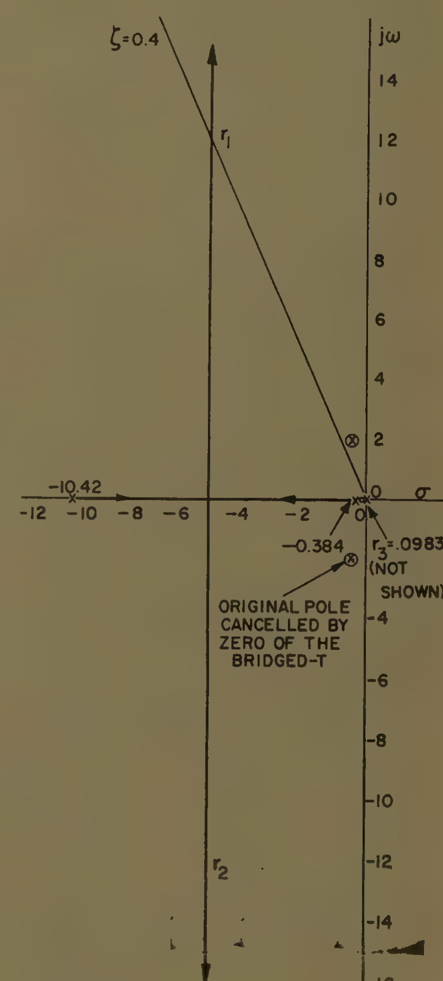


Fig. 22. Root loci of $\dfrac{40K(s+0.1)}{s(s+0.304)(s+10.42)}$

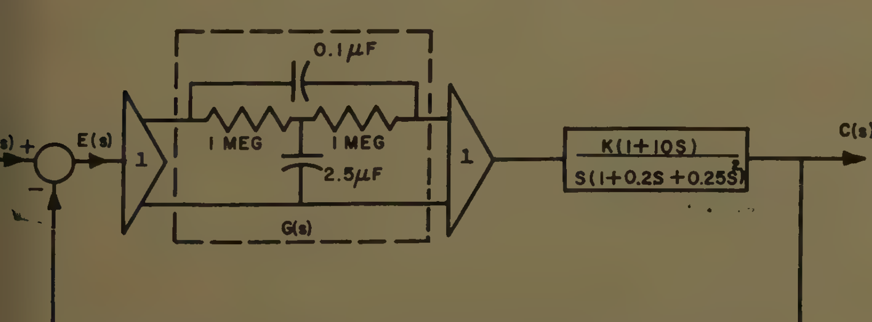


Fig. 21. Control system with bridged-T as compensating network G(s)

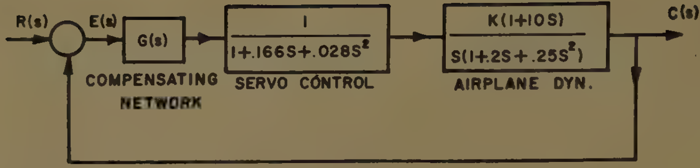


Fig. 23. Block diagram of the system for example 2

With a knowledge of the root locus technique and the results of the analysis given in the last two sections, the proper type of bridged-T can be selected to meet system requirements. Generally, the effect of loading can be used to adjust the pole positions of the transfer function of the bridged-T. However, while it is impossible to fix both poles at will, one can fix the one pole which is most significant. The capacitive load for bridged-T type *Ib* is used for pole adjustment in the numerical example 2 in the Appendix.

In actual applications, the bridged-T without load should be designed first by consultation of the design charts. Zeros of this network are fixed at the desired positions; then the complete pole-zero configuration of $G(s)$ is obtained. Modification of $G(s)$ by loading effects is made by selecting the proper type of loading. The exact pole-zero location of $G'(s)$ is obtained by applying the root locus technique to the equivalent control system $Z'(s)/Z(s)$.

Application

With the developed design chart, one can easily place a pair of conjugate complex zeros at any point in the complex plane. However, their associated unloaded poles cannot be controlled. For optimum compensation it is often desirable to move the bridged-T poles. This can be done by connecting the proper load impedance at the output terminals and using root locus techniques to determine the new pole-zero configuration. With this method, only one pole can be fixed in a specified position. In many cases, however, there will be one pole that is more significant than the other, and this technique may still be used.

Probably the most common application of bridged-T networks to d-c servo systems would be in those systems in which the transfer functions contain complex poles. The simplest way to compensate these types would be use of the cancellation technique of putting zeros on top of the complex poles. The system to be employed for illustration is the airplane auto pilot. Two examples are given in the Appendix. In example 1, only the bridged-T without loading effect will be used, but in example 2, the loading effect will be applied as well.

It should be stated that these examples are mainly intended to illustrate techniques for improving the system stability. If all specifications of the system are not entirely met, additional measures would, of course, be used. It should also be recognized that in actual systems, the exact pole-zero locations may not be exactly known and will often vary with change in environmental conditions. Exact cancellation of terms is, therefore, seldom if ever achieved.

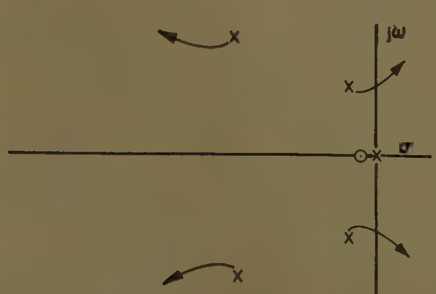


Fig. 24. Sketch of root loci of equation 35

Appendix

Numerical Example 1

The control system shown in Fig. 19 is selected to demonstrate the use of a bridged-T network. $G(s)$ is the compensating bridged-T network. The open-loop control transfer function, in this case, represents the short-period longitudinal mode of an airplane. The fin control servo is assumed perfect to avoid its effects on system performance. Without compensation, or $G(s)=1$, the root loci of this system are shown in Fig. 20. Without compensation, because of the small damping ratio, the system transient overshoot will be high.

With Bridged-T as Compensating Network

Pole cancellation will be used to improve the stability of the system. The complex poles in the original transfer function have a damping ratio of 0.2 and a natural frequency of 2 radians per second. In applying the cancellation method, the bridged-T network is designed to give complex zeros at $\zeta=0.2$ and $\omega_n=2$. Generally, either type of bridged-T can be used. For this particular case, the bridged-T type *Ib* has already been synthesized in the previous example 2 and will be used as the compensating network $G(s)$. The complete diagram of this control system is shown in Fig. 21. From the previous result

$$G(s)=\frac{(s+0.4+j1.95)(s+0.4-j1.95)}{(s+0.384)(s+10.42)}$$

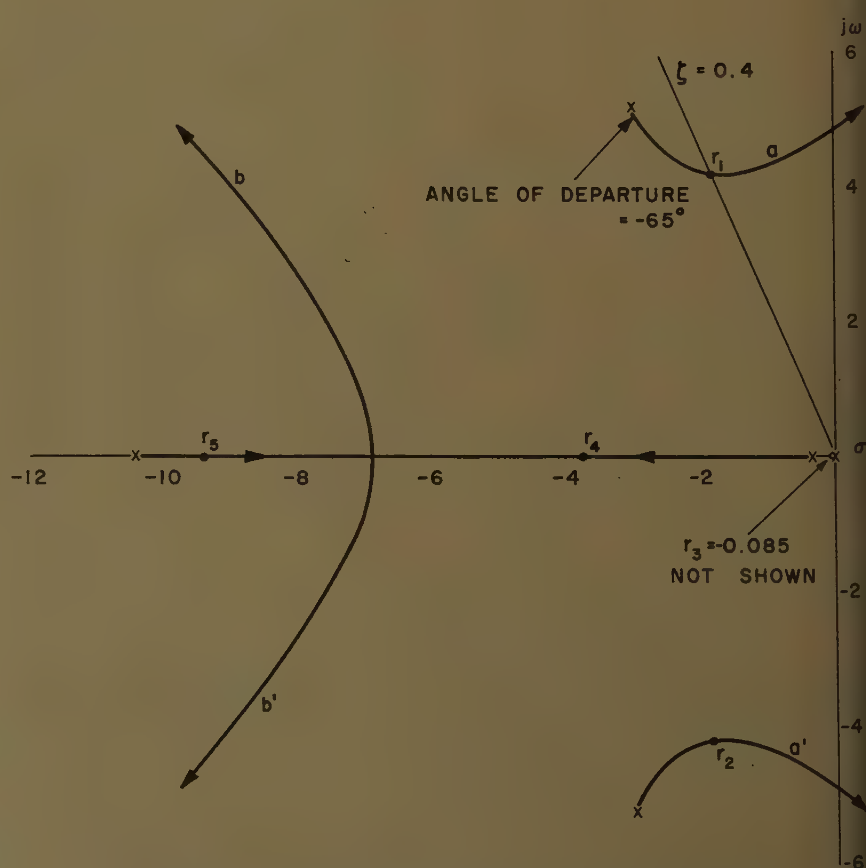
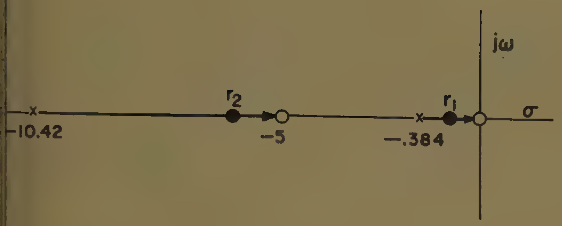


Fig. 25. Root loci of the system compensated by bridged-T without load



$$)=\frac{(1+0.2s+0.25s^2)}{(1+2.7s)(1+0.096s)} \qquad (31)$$

$$)=\frac{K(1+10s)}{s(1+2.7s)(s+10.42)} \qquad (32)$$

re $F_0(s)$ is the new open-loop transfer ction, or

$$)=\frac{40K(s+0.1)}{s(s+0.384)(s+10.42)} \qquad (33)$$

ot loci of this function are shown in Fig.

n this particular case, with a bridged-T compensating network, the new root loci w more satisfactory results. The loci nearly vertical, as the zero at -0.1 and pole at the origin are very close together en compared with the other singulari- .

th $\zeta=0.4$

4.28

$_2=-5.24\pm j12.1$

-0.0983

Vith $r(t)=$unit step input, residue at

$0.537\,\underline{|156°}$

$0.537\,\underline{|-156°}$

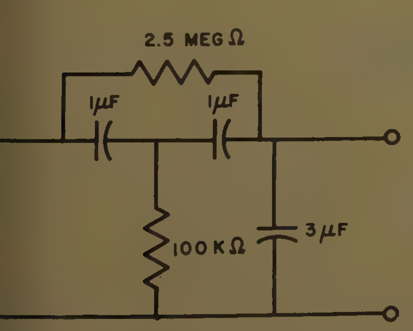
-0.0172

en

$$=1-0.0172\epsilon^{-0.0983t}+1.074\epsilon^{-5.24t}\cos(12t+156°) \qquad (34)$$

Vith $\zeta=0.4$, the overshoot from r_1 and r_2 5%². In this case, the residue of the t r_3 (with unit step input) is nega- and very small, so the overshoot will slightly less than 25%. Actually, due he very long time constant of r_3, it will a long-tail effect to the closed loop performance. However, the residue at this root is so small that, in practice, one can neglect it.

It should be pointed out that this analysis is made on the basis of exact cancellation which, of course, is impossible. The technique can still be applied for if the zeros are still near the poles, the residues at those roots will be very small and they will not effect the closed-loop performance significantly.



27. Compensating network for example 2

Fig. 26. Sketch of root loci for poles movement versus C_t

Numerical Example 2

The application of bridged-T with loading effect will be illustrated in this example to improve the system performance. The same problem used in example 1 is selected for illustration but, in addition, another servo control unit is placed in cascade with the original. This system is shown in Fig. 23.

With $G(s)=1$

$$F_0(s)=\frac{K(1+10s)}{s(1+0.25+0.25s^2)\times(1+0.166s+0.028s^2)} \qquad (35)$$

The root loci of this system are shown in Fig. 24. Without compensation, it is seen that the predominating roots of the system are about to move to the right half-plane, even at the very low gain. It is desired to design a compensating network $G(s)$ to improve the system stability. For this particular problem, the bridged-T networks again seem to to be very applicable.

Bridged-T Applied Without Load

If the bridged-T network, which has been used in the last problem, is applied to this system as a compensating network $G(s)$ to cancel the complex poles of the airplane function, the system stability is improved. See Fig. 25 for the system root loci plot.

If the damping ratio of 0.4 is selected for gain adjustment, the roots of the closed-loop system are

$r_1, r_2=-1.75\pm j4.15$

$r_3=-0.085$

$r_4=-3.7$

$r_5=-9.39$

The open-loop gain K equals 0.446. The most undesirable features are the large residue of r_3, which is close to the origin and the low gain. When a unit step input is applied, the transient term caused by this root is $-0.162\epsilon^{-0.085t}$.

Due to this large residue and the very long time constant of about 11.8 seconds, the closed-loop response to a unit step input will have a long tail, and the system performance is not considered satisfactory.

Loading Effect Applied for Poles Adjustment

The root loci plot in Fig. 25 shows that the closed-loop response can be improved by moving both poles at -0.384 and -10.42 towards the origin until the loci a and a' are about to change to the left (or b and b' are about to change to the right). The required location of the new poles can be roughly estimated by checking the angle of departure from the complex poles. If all singularities are held constant, the pole at -10.42 has to be moved to -6 to satisfy this condition. If a capacitance is used for loading the original bridged-T, both poles at -10.42 and -0.384 can be moved to the right.

Another improvement that may be considered at the same time is to try to cancel the zero at -0.1 by moving the pole from -0.384. If an approximate cancellation at this point can be made, the calculation is simplified and the effect of the long time constant transient term will be eliminated.

For this particular problem, the loading effect of case *IIb* seems very suitable. The bridged-T type *Ib* used with load capacitance will have terminating loci from both poles at -5 and the origin. If the case *IIa* is used, the loci will terminate at -0.8 and origin. In case *IIb*, there is a possibility of obtaining both zero cancellation and another pole moving to the vicinity of -6, while, in case *IIa*, the pole at -0.384 will change only slightly. The previous results from the analysis of loading case *IIb* will be applied for the adjustment of the poles of bridged-T type *Ib*.

First, the bridged-T type *Ib* (without load) containing zeros at the same location of complex poles in the airplane transfer function is designed: $\zeta_{nz}=0.2$; $1/R_1C=10$; $R_2/R_1=25$.

If $C=1\mu$F is selected, then $R_1=0.1$ megohm and $R_2=2.5$ megohms. The pole-zero configuration of the voltage transfer function is still the same as for the network *Ia* which was used in the last example.

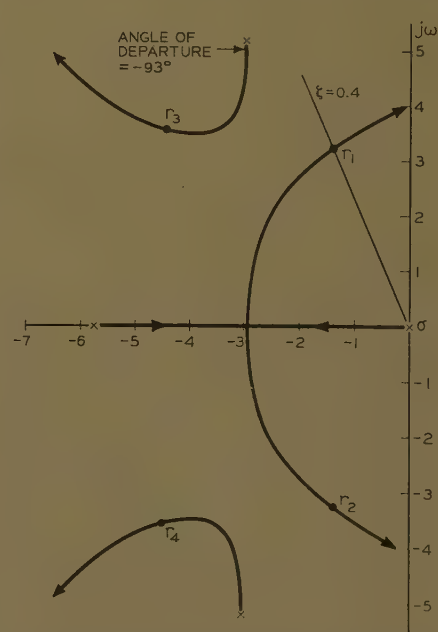


Fig. 28. Root loci of the system compensated by bridged-T type Ib with capacitive load

Next, using the capacitive load C_t and applying equation 27 to obtain the relation for pole adjustment yields

$$\frac{2C_t s(s+5)}{(s+10.42)(s+0.384)} = -1 \qquad (36)$$

See Fig. 26 for the root loci of pole movement.

If the root r_1 is fixed at -0.1, then $C_t = 2.95$, and $r_2 = -5.75$. The network which is to be used for $G(s)$ is shown in Fig. 27. The root loci plot of the system is given in Fig. 28. With the damping ratio $= 0.4$, the results are $r_1, r_2 = -1.4 \pm j3.25$; $r_3, r_4 = -4.5 \pm j3.55$; $K = 2.04$.

If $r(t) =$ unit step input

$$c(t) = 1 + 1.576\epsilon^{-1.4t}\cos(3.25t + 96°) + 0.89\epsilon^{-4.5t}\cos(3.55t + 200°) \qquad (37)$$

The overshoot in this case is roughly estimated at about 26%.

References

1. Control-System Dynamics (book), W. R. Evans. McGraw-Hill Book Company, Inc., New York, N. Y., 1954, pp. 156–57.

2. Automatic Feedback Control System Synthesis (book), John G. Truxal. McGraw-Hill Book Company, Inc., 1955, pp. 38–43.

Discussion

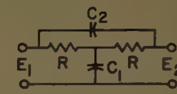
Theodore A. Savo (Hughes Aircraft Company, Culver City, Calif.): The authors are to be complimented on their interesting paper, which helps to bridge the gap between what might be called network synthesis and network design. The root-locus method, used in the graphical study of the poles and zeros of the bridged-T network in this paper, can of course be used for other networks the extension being straightforward. If the decision has been made to compensate in the "open-loop" sense and the characteristics of the bridged-T are satisfactory, then the content of this paper provides useful information for the rapid design of the network with and without loading. An interesting feature is that the design is affected in terms of the poles and zeros of the voltage transfer function for the network, and hence allows the network design to be combined with a root-locus study of the control system characteristics.

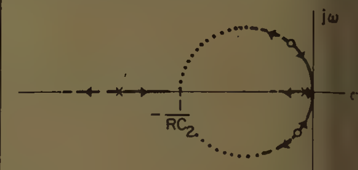
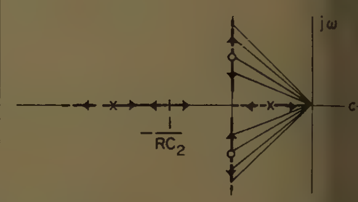
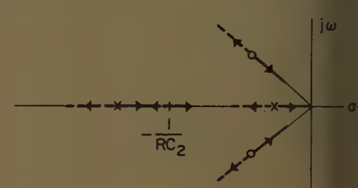
The purpose of this discussion is to provide some parallel opinions on the subject matter, to offer some extensions to the content, and to indicate cases where the method has proved useful. The authors have adequately studied the characteristics of the bridged-T (types *Ia* and *Ib*) from a practical-design viewpoint. One additional consideration is the problem of parameter variations. Table I(A) and (B) conveniently summarizes the effects of component changes on the voltage transfer function $E_2/E_1(s)$.

It is interesting to note that the study of the effects of a load on the transfer characteristics of the bridged-T can be conveniently expressed in terms of the short-circuit admittance parameters (y_{12}, y_{22}).

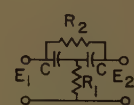
Table I. Summary of Network Characteristics for Parameter Variations

A. NETWORK TYPE Ia: $\zeta = \sqrt{\frac{C_2}{C_1}}$; $\omega_n = \frac{1}{R\sqrt{C_1 C_2}}$



PARAMETER	VARIATION	CHARACTERISTIC CHANGE	POLE–ZERO EFFECTS: E_2/E_1
C_1	INCREASED	$\frac{C_1}{C_2}$ — INCREASES ζ — DECREASES $\frac{1}{RC_2}$ — UNCHANGED	
	DECREASED	$\frac{C_1}{C_2}$ — DECREASES ζ — INCREASES $\frac{1}{RC_2}$ — UNCHANGED	
C_2	INCREASED	$\frac{C_1}{C_2}$ — DECREASES ζ — INCREASES $\frac{1}{RC_2}$ — DECREASES (FASTER THAN ζ INCREASES)	
	DECREASED	$\frac{C_1}{C_2}$ — INCREASES ζ — DECREASES $\frac{1}{RC_2}$ — INCREASES (FASTER THAN ζ DECREASES)	
R	BOTH INCREASED SIMULTANEOUSLY	$\frac{C_1}{C_2}$ — UNCHANGED ζ — CONSTANT $\frac{1}{RC_2}$ — DECREASES	
	BOTH DECREASED SIMULTANEOUSLY	$\frac{C_1}{C_2}$ — UNCHANGED ζ — CONSTANT $\frac{1}{RC_2}$ — INCREASES	

B. NETWORK TYPE Ib: $\zeta = \sqrt{\frac{R_1}{R_2}}$; $\omega_n = \frac{1}{C\sqrt{R_1 R_2}}$



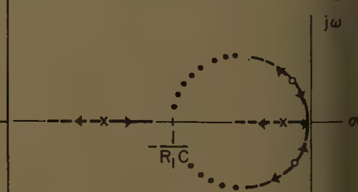
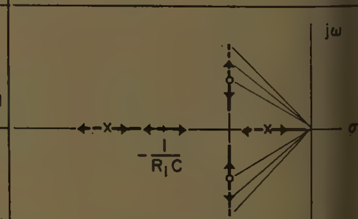
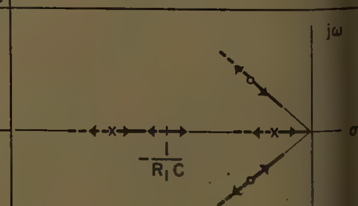
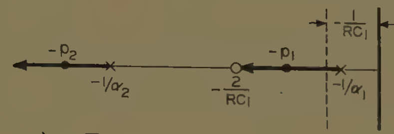
PARAMETER	VARIATION	CHARACTERISTIC CHANGE	POLE–ZERO EFFECTS: E_2/E_1
R_2	INCREASED	$\frac{R_2}{R_1}$ – INCREASES ζ – DECREASES $\frac{1}{R_1 C}$ – UNCHANGED	
	DECREASED	$\frac{R_2}{R_1}$ – DECREASES ζ – INCREASES $\frac{1}{R_1 C}$ – UNCHANGED	
R_1	INCREASED	$\frac{R_2}{R_1}$ – DECREASES ζ – INCREASES $\frac{1}{R_1 C}$ – DECREASES (FASTER THAN ζ INCREASES)	
	DECREASED	$\frac{R_2}{R_1}$ – INCREASES ζ – DECREASES $\frac{1}{R_1 C}$ – INCREASES (FASTER THAN ζ DECREASES)	
C	BOTH INCREASED SIMULTANEOUSLY	$\frac{R_2}{R_1}$ – UNCHANGED ζ CONSTANT $\frac{1}{R_1 C}$ – DECREASES	
	BOTH DECREASED SIMULTANEOUSLY	$\frac{R_2}{R_1}$ – UNCHANGED ζ – CONSTANT $\frac{1}{R_1 C}$ – INCREASES	

Table II. Summary of Loaded Types Ia and Ib Networks

(A) Network Ia

$Y_L=1/R_t$

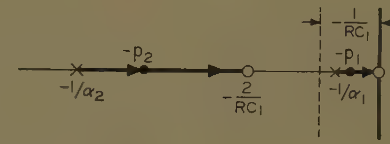


$$\frac{Y_L}{y_{22}}(S)=\frac{2R}{R_t}\left[\frac{\left(1+\frac{RC_1}{2}S\right)}{(1+\alpha_1S)(1+\alpha_2S)}\right]$$

$$G(S)=\frac{1+aS+bS^2}{(1+\alpha_1S)(1+\alpha_2S)}$$

$$G'(S)=\frac{R_t}{2R+R_t}\left[\frac{1+aS+bS^2}{\left(1+\frac{S}{p_1}\right)\left(1+\frac{S}{p_2}\right)}\right]$$

(b) $Y_L=C_tS$

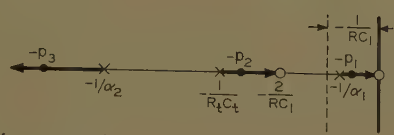


$$\frac{Y_L}{y_{22}}(S)=2RC_t\left[\frac{S\left(1+\frac{RC_1}{2}S\right)}{(1+\alpha_1S)(1+\alpha_2S)}\right]$$

$$G(S)=\frac{1+aS+bS^2}{(1+\alpha_1S)(1+\alpha_2S)}$$

$$G'(S)=\frac{1+aS+bS^2}{\left(1+\frac{S}{p_1}\right)\left(1+\frac{S}{p_2}\right)}$$

$$Y_L=\left[\frac{C_tS}{R_tC_tS+1}\right]$$

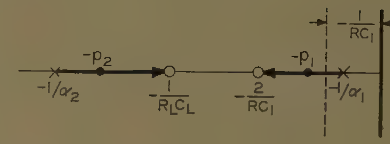


$$\frac{Y_L}{y_{22}}(S)=\frac{2RC_tS\left(1+\frac{RC_1}{2}S\right)}{(1+\alpha_1S)(1+\alpha_2S)(1+R_tC_tS)}$$

$$G(S)=\frac{1+aS+bS^2}{(1+\alpha_1S)(1+\alpha_2S)}$$

$$G'(S)=\frac{(1+aS+bS^2)(1+R_tC_tS)}{\left(1+\frac{S}{p_1}\right)\left(1+\frac{S}{p_2}\right)\left(1+\frac{S}{p_3}\right)}$$

(d) $Y_L=\left[\frac{R_LC_LS+1}{R_L}\right]$



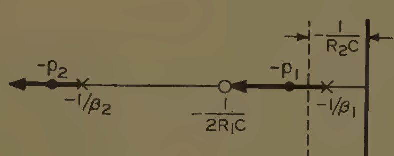
$$\frac{Y_L}{y_{22}}(S)=\frac{2R}{R_L}\left[\frac{\left(1+\frac{RC_1}{2}S\right)(1+R_LC_LS)}{(1+\alpha_1S)(1+\alpha_2S)}\right]$$

$$G(S)=\frac{1+aS+bS^2}{(1+\alpha_1S)(1+\alpha_2S)}$$

$$G'(S)=\frac{R_L}{R_L+2R}\left[\frac{1+aS+bS^2}{\left(1+\frac{S}{p_1}\right)\left(1+\frac{S}{p_2}\right)}\right]$$

(B) Network Ib

$Y_L=1/R_t$

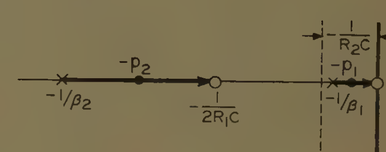


$$\frac{Y_L}{y_{22}}(S)=\frac{R_2}{R_t}\left[\frac{(1+2R_1CS)}{(1+\beta_1S)(1+\beta_2S)}\right]$$

$$G(S)=\frac{1+\bar{a}S+\bar{b}S^2}{(1+\beta_1S)(1+\beta_2S)}$$

$$G'(S)=\frac{R_t}{R_2+R_t}\left[\frac{1+\bar{a}S+\bar{b}S^2}{\left(1+\frac{S}{p_1}\right)\left(1+\frac{S}{p_2}\right)}\right]$$

(b) $Y_L=C_tS$

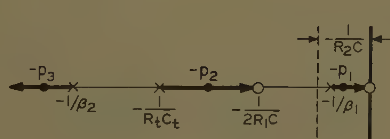


$$\frac{Y_L}{y_{22}}(S)=\frac{R_2C_tS(1+2R_1CS)}{(1+\beta_1S)(1+\beta_2S)}$$

$$G(S)=\frac{1+\bar{a}S+\bar{b}S^2}{(1+\beta_1S)(1+\beta_2S)}$$

$$G'(S)=\frac{1+\bar{a}S+\bar{b}S^2}{\left(1+\frac{S}{p_1}\right)\left(1+\frac{S}{p_2}\right)}$$

$$Y_L=\left[\frac{C_tS}{1+R_tC_tS}\right]$$

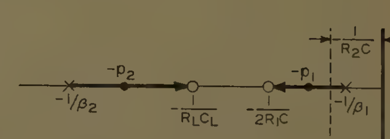


$$\frac{Y_L}{y_{22}}(S)=\frac{R_2C_tS(1+2R_1CS)}{(1+\beta_1S)(1+\beta_2S)(1+R_tC_tS)}$$

$$G(S)=\frac{1+\bar{a}S+\bar{b}S^2}{(1+\beta_1S)(1+\beta_2S)}$$

$$G'(S)=\frac{(1+\bar{a}S+\bar{b}S^2)(1+R_tC_tS)}{\left(1+\frac{S}{p_1}\right)\left(1+\frac{S}{p_2}\right)\left(1+\frac{S}{p_3}\right)}$$

(d) $Y_L=\left[\frac{R_LC_LS+1}{R_L}\right]$



$$\frac{Y_L}{y_{22}}(S)=\frac{R_2}{R_L}\left[\frac{(R_LC_LS+1)(1+2R_1CS)}{(1+\beta_1S)(1+\beta_2S)}\right]$$

$$G(S)=\frac{1+\bar{a}S+\bar{b}S^2}{(1+\beta_1S)(1+\beta_2S)}$$

$$G'(S)=\frac{R_L}{R_2+R_L}\left[\frac{1+\bar{a}S+\bar{b}S^2}{\left(1+\frac{S}{p_1}\right)\left(1+\frac{S}{p_2}\right)}\right]$$

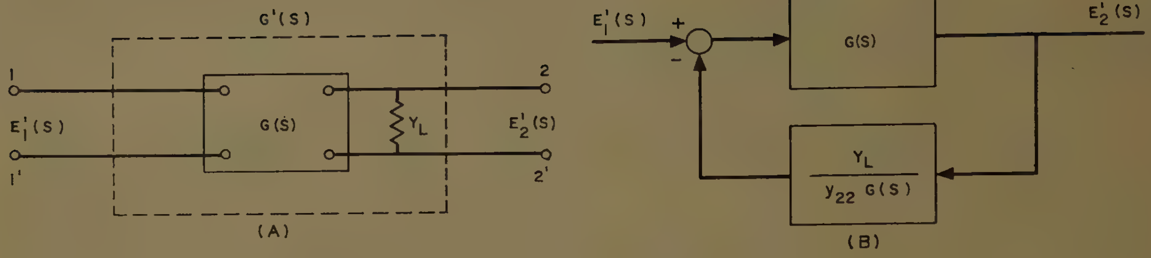


Fig. 29. Netwo loading and equivi lent feedback syste

The voltage transfer function for the network with no load is

$$\frac{E_2}{E_1}(S)=\frac{y_{12}}{y_{22}}\triangleq G(S) \tag{38}$$

(Since it would add little to the discussion, the minus sign normally associated with equation 38 has been arbitrarily omitted.)

The expression for the voltage transfer function if the network is loaded with an admittance Y_L can be expressed as

$$\frac{E_2}{E_1}(S)=\frac{y_{12}}{y_{22}+Y_L}\triangleq G'(S) \tag{39}$$

Slight algebraic manipulation of equation 39 yields

$$G'(S)=\frac{E_2}{E_1}(S)=\frac{y_{12}/y_{22}}{1+\left[\frac{Y_L}{y_{22}}\left(\frac{y_{22}}{y_{12}}\right)\right]\left(\frac{y_{12}}{y_{22}}\right)}\triangleq$$

$$\frac{G(S)}{1+\left(\frac{Y_L}{y_{22}G(S)}\right)G(S)} \tag{40}$$

Equation 40 defines an equivalent feedback control system in which the closed loop characteristics represent the transfer function of the network with load; see Fig. 29. The open-loop gain from Fig. 29 and equation 40 is

$$\text{Loop gain}=-\left(\frac{Y_L}{y_{22}}\right) \tag{41}$$

Equation 41 can be correlated with the method used in the paper, namely

$$\frac{Y_L}{y_{22}}\triangleq\frac{Z'(S)}{Z(S)} \tag{42}$$

The characteristics of the loaded network $G'(s)$ can now be studied by the root-locus method, the characteristic equation being

$$1+\frac{Y_L}{y_{22}}(S)=0 \tag{43}$$

where y_{22} = short-circuit admittance function for the unloaded network $G(s)$; and Y_L = admittance of any arbitrary load. Specifically, the short-circuit admittance functions for the two types of bridged-T networks are, for type *Ia:*

$$y_{22}=\frac{1+R(C_1+2C_2)S+C_1C_2R^2S^2}{2R\left(1+\frac{RC_1}{2}S\right)} \tag{44}$$

and for type *Ib:*

$$y_{22}=\frac{1+C(R_2+2R_1)S+R_1R_2C^2S^2}{R_2(1+2R_1CS)} \tag{45}$$

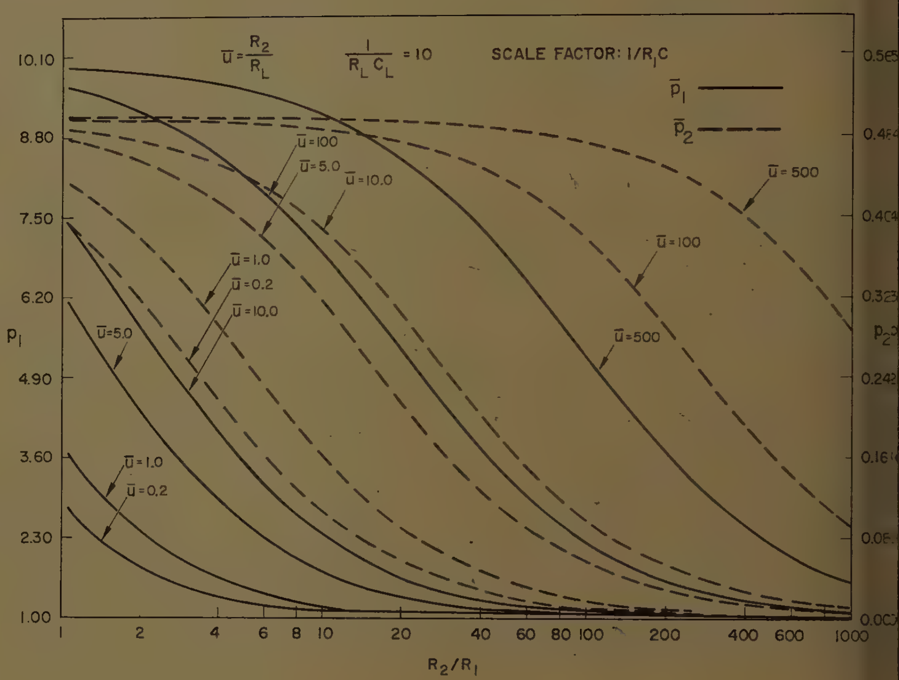


Fig. 30. Type Ib bridged-T pole positions with load $Y_L=(R_LC_L+1)/R_L$

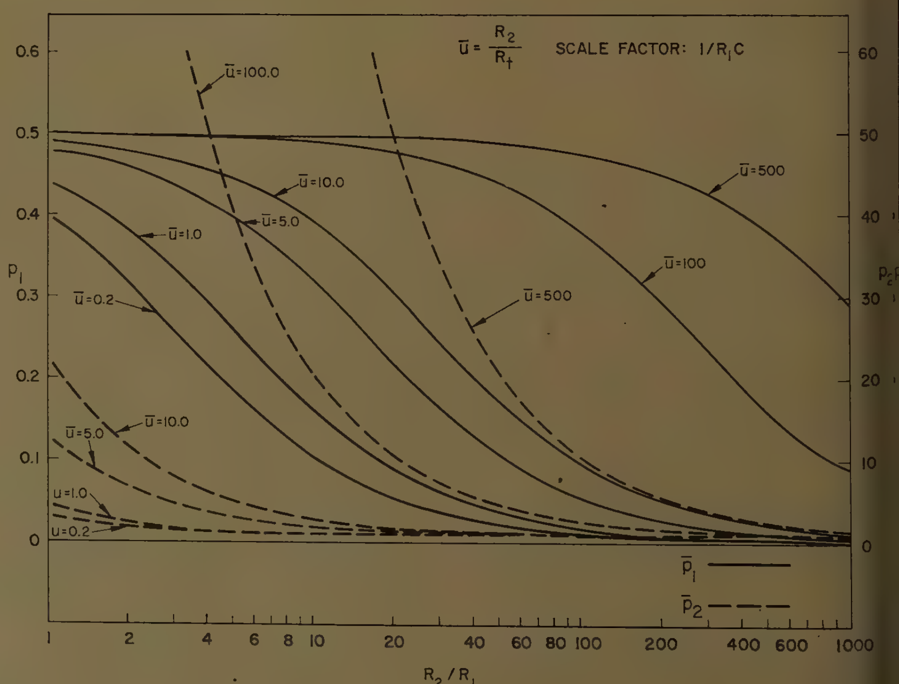


Fig. 31. Type Ib bridged-T pole positions with load $Y_L=1/R_t$

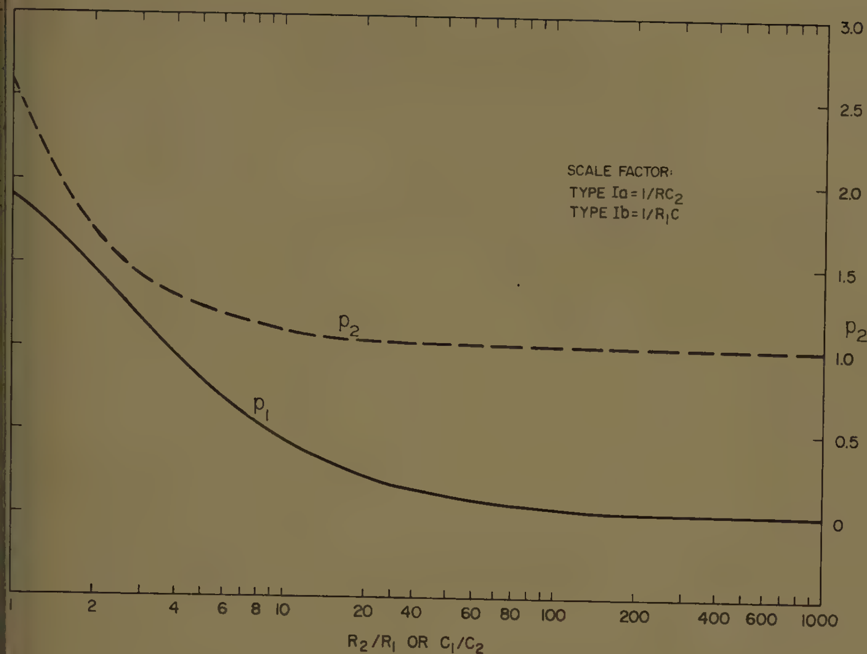


Fig. 32. Bridged-T pole positions, unloaded

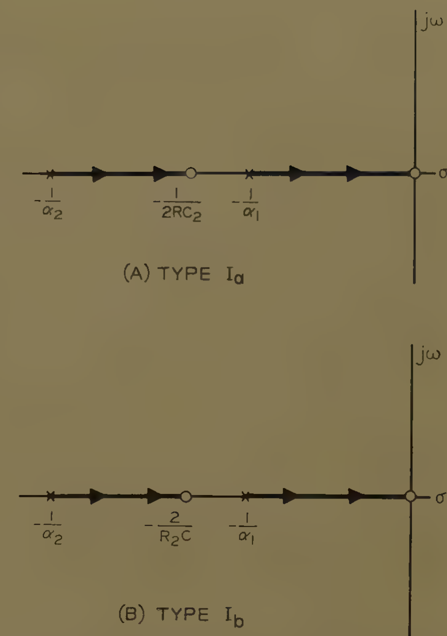


Fig. 34. Bridged-T pole locations with resistive source load Rs

ne authors have studied the bridged-T acteristics with three types of simple s. One additional type of simple load, considered, is the case of a parallel comtion of one resistor and capacitor. this case the load admittance is

$$\frac{R_L C_L S+1}{R_L} \tag{46}$$

function $Y_{L/y_{22}}$ becomes, for type *Ia:*

$$S) \triangleq \frac{Z'(S)}{Z(S)} = \frac{2R}{R_L} \frac{\left(1+\frac{RC_1}{2}S\right)(1+R_L C_L S)}{(1+\alpha_1 S)(1+\alpha_2 S)} \tag{47}$$

for type *Ib:*

$$S) \triangleq \frac{Z'(S)}{Z(S)} = \frac{R_2}{R_L} \frac{(1+2R_1 CS)(1+R_L C_L S)}{(1+\beta_1 S)(1+\beta_2 S)} \tag{48}$$

e zeros of the transfer function with arallel R_L, C_L load, $G'(S)$, are, of course, same as the zeros of the network fer function without load $G(S)$. Both orks, however, have a d-c attenuation. onvenient summary of the loading s on the pole locations is given in Table) and (B) for the three cases considered e paper and for the additional case just nted. The information shown in Table useful in determining qualitatively ffects of loading (intentional or unintentional); however, in a specific problem, if the pole locations are to be determined exactly, a design chart of the form shown in Fig. 30 is appropriate for engineers interested in rapid laboratory-type design. Fig. 30 depicts the type *Ib* network with the parallel, R_L, C_L type of load and the normalized time constant of the load set at $T_L=0.1$ second. Fig. 31 illustrates the pole positions for the type *Ib* network with simple resistive loading R_t. Fig. 32 gives the pole positions for the unloaded bridged-T. It should be noted that the pole positions shown in Fig. 30 and 31 are computed with R_1C normalized to unity; hence a $1/R_1C$ scale factor is necessary. Similar comments apply to Fig. 32.

It is logical to consider the characteristics of the bridged-T network for the case of source loading as illustrated in Fig. 33.

The transfer function for the case of source loading is, for type *Ia:*

$$\bar{G}(S) = \frac{1+2RC_2S+C_1C_2R^2S^2}{1+[R(C_1+2C_2)S+R_sC_1]S+RC_1C_2(R+2R_s)S^2} \tag{49}$$

and for type *Ib:*

$$\bar{G}(S) = \frac{1+2R_1CS+R_1R_2C^2S^2}{1+C(2R_1+R_2+2R_s)S+R_2(R_1+R_s)C^2S^2} \tag{50}$$

As might be expected for resistive source loading, equations 49 and 50 reveal that the zeros of the network are not altered. Study of the pole migration using the root-locus method is best effected in the case of source loading by using the open-circuit impedance functions $z_{12}(S)$ and $z_{11}(S)$. The voltage transfer function of the network with no load is

$$\frac{E_2}{E_1}(S) = \frac{z_{12}(S)}{z_{11}(S)} \tag{51}$$

If an arbitrary source load Z_s is added, then the voltage transfer function becomes

$$\frac{E_2}{E_1}(S) = \frac{z_{12}}{z_{11}+Z_s} \tag{52}$$

Manipulation of equation 52, in a manner analogous to equation 40, yields

$$\frac{E_2}{E_1}(S) = \frac{z_{12}/z_{11}}{1+\frac{Z_s}{z_{11}}} \tag{53}$$

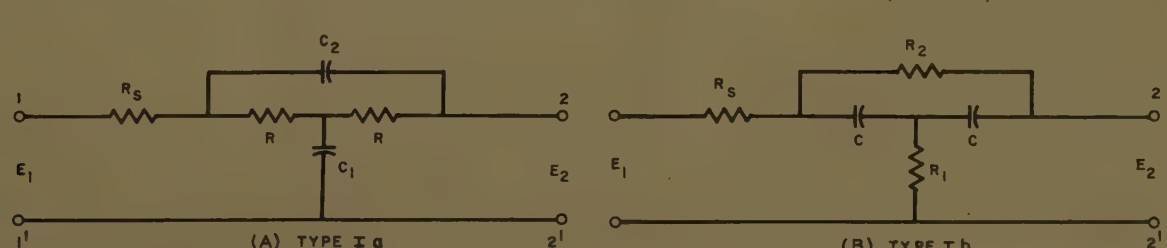
Equation 53 indicates that the poles of the transfer characteristic can be studied with a root-locus plot of Z_s/z_{11} in a manner analogous to Y_L/y_{22} for load impedances: see equation 43.

The open-circuit driving point impedance, $z_{11}(S)$, for the bridged-T network is, for type *Ia:*

$$z_{11} = \frac{1+R(C_1+2C_2)S+C_1C_2R^2S^2}{C_1S(1+2RC_2S)} \tag{54}$$

and for type *Ib:*

$$z_{11} = \frac{1+C(2R_1+R_2)S+R_1R_2C^2S^2}{2CS\left(1+\frac{R_2C}{2}S\right)} \tag{55}$$



3. Bridged-T rk with source loading

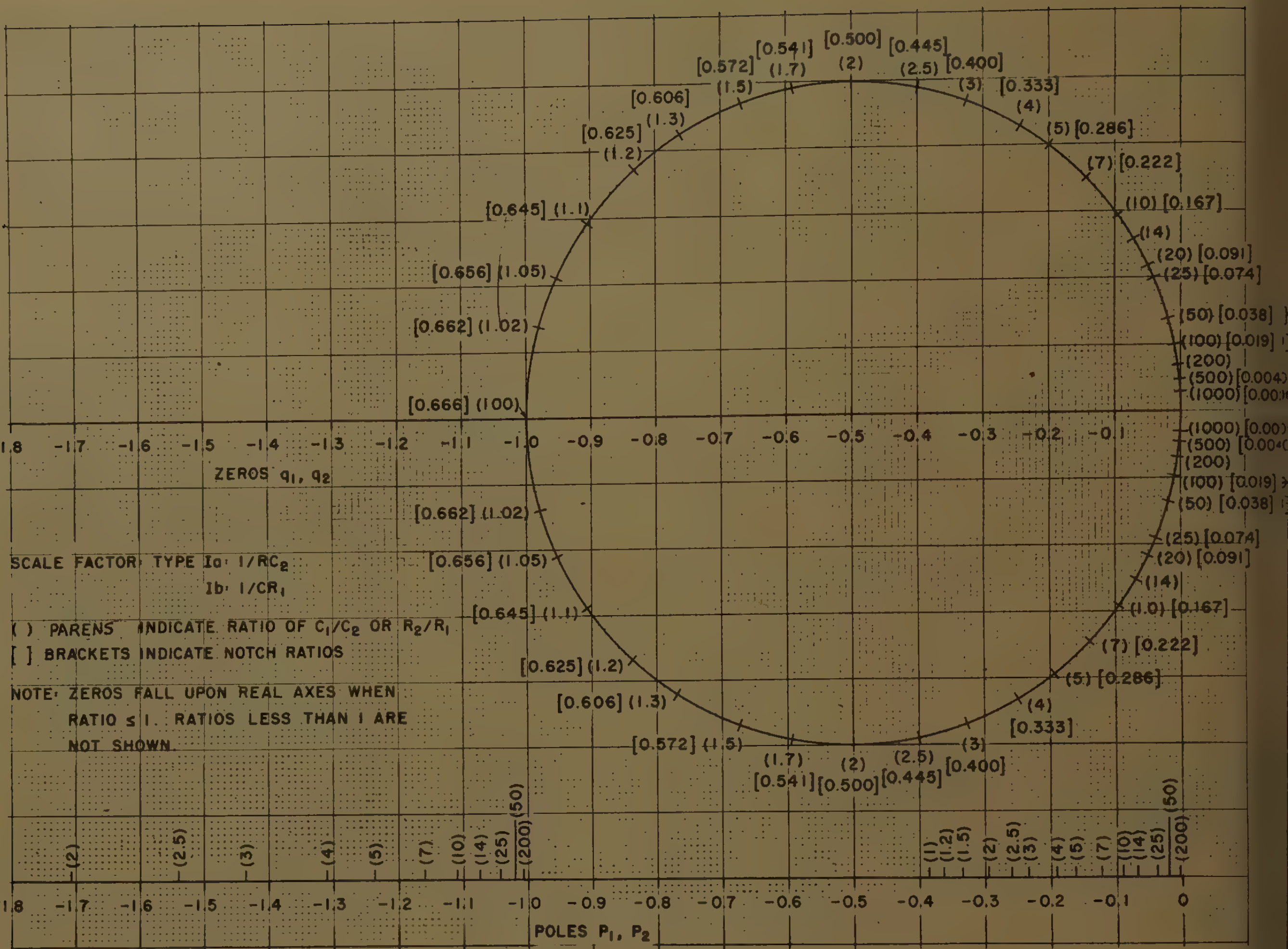


Fig. 35. Design chart for bridged-T type Ia and Ib taken from normalized equation 5

$$\frac{E_{out}}{E_{in}}=\frac{1+2s+as^2}{1+(2+a)s+as^2}$$

The resultant ratio $Z_s/z_{11}(S)$ is $(Z_s=R_s)$ for type *Ia:*

$$\frac{Z_s}{z_{11}}(S)=\frac{R_sC_1S(1+2RC_2S)}{(1+\alpha_1S)(1+\alpha_2S)} \tag{56}$$

and for type *Ib:*

$$\frac{Z_s}{z_{11}}(S)=\frac{2R_sCS\left(1+\frac{R_2C}{2}S\right)}{(1+\alpha_1S)(1+\alpha_2S)} \tag{57}$$

The parameters $1/\alpha_1$ and $1/\alpha_2$ in equations 56 and 57 characterize the pole locations for $R_s=0$.

Figure 34 depicts the pole migration derived from equations 56 and 57. With the frequency normalization used for the unloaded bridged-T, the gain calibration of the loci of Fig. 34 is quite simple, for example, for type *Ia:*

$$RC_2=1$$

$$C_1/C_2=n$$

$$R_sC_1=n\left(\frac{R_s}{R}\right) \tag{58}$$

and for type *Ib:*

$$R_1C=1$$

$$\frac{R_2}{R_1}=n$$

$$R_sC=n\left(\frac{R_s}{R_2}\right) \tag{59}$$

Consequently, for a given n value which fixes the complex zeros, the variation in the pole locations can be studied by varying the ratios (R_s/R) and $R_s/R_2)$ respectively. In conclusion, it is obvious that if the network level can be designed to yield $R>>R_s$ or $R_2>>R_s$, then the unloaded design charts can be used. In cases where the source resistance is not negligible or drifts substantially, the results shown in Fig. 34 are useful.

The case of both input and output loading has also been considered by the aforementioned method; however, the number of variable parameters becomes large unless special symmetry is assumed.

In conclusion it is worth mentioning that the root-locus method of designing the bridged-T has been successfully used in effecting the compensation of both electric and hydraulic type of feedback systems. Cancellation-type compensation using the bridged-T network is particularly useful in laboratory-type design and testing when the characteristics of the "fixed" part of the system are not known exactly or when they vary because of nonlinearities. It should also be mentioned that a simple lead network in tandem with the bridged-T can be used to effectively move the low-frequency pole of the unloaded bridged-T to a higher frequency.

Propat Chandaket and Allen B. Rosenste[in]: The authors wish to thank Dr. Savo [for] his interesting discussion and his additi[onal contribution] to the work of this paper. The use [of] short-circuit admittance parameters [to] study the loading effect is more convenie[nt] for the circuit designers. Dr. Savo's pl[ots] of pole positions for the case of bridged[-T] under no load and under loaded conditi[ons] of pure resistance and parallel R-C will [be] particularly useful for the designer.

Since the bridged-T is frequently e[m]ployed as an a-c filter, the authors wo[uld] like to extend the utility of the method [de]veloped a little further by adding the not[ch] ratio and notch width characteristics of t[he] unloaded bridged-T of both types.

These properties can be obtained qu[ite] easily from the normalized equation in F[ig.] 32.

$$\frac{E_{out}}{E_{in}}=\frac{1+2s+as^2}{1+(2+a)s+as^2}=\frac{\left(1+\frac{s}{q_1}\right)\left(1+\frac{s}{q_2}\right)}{\left(1+\frac{s}{p_1}\right)\left(1+\frac{s}{p_2}\right)}$$

where a = ratio of C_1/C_2 (type *Ia*) or R_2/R_1 (type *Ib*).

We note from the equation that the pr[od]uct of the zeros equals the product of [the] poles, i.e., $q_1q_2=p_1p_2=1/a$.

f the zeros are complex, then their malized natural frequency is $\omega_{n2} 1/\sqrt{a}$; equation 8(A). This, of course, is the quency in radians locating the center (in og scale) of the notch, i.e., the double ak point of the zeros.

The d-c gain of the bridged-T is unity. The notch ratio is defined as the ratio of n at the notch frequency to the gain at o frequency, replacing s with $j(1/\sqrt{a})$ following is obtained from equation 1:

$$\text{ch ratio} = \frac{2}{2+a} \tag{61}$$

This shows that the notch ratio depends only on the ratio of C_1/C_2 or R_2/R_1.

For design convenience, the design chart shown in Fig. 3 has been redrawn with the notch ratios for the various values of C_1/C_2 or R_2/R_1 added in brackets; see Fig. 35.

The notch width frequency is defined as the frequency difference between the two poles of the bridged-T.

From equation 60

$$p_1 = -\frac{(2+a)}{2a} + \frac{\sqrt{4+a^2}}{2a}$$

$$p_2 = -\frac{(2+a)}{2a} - \frac{\sqrt{4+a^2}}{2a}$$

This gives

$$\text{Normalized notch width} = \frac{\sqrt{4+a^2}}{a} \tag{62}$$

To obtain the real value of notch width frequency, one must multiply the result by its scale factor. Since $p_1 p_2 = 1/a$ and the zeros fall at $1/\sqrt{a}$, the pole-zero configuration on a log scale such as in a Bode diagram is symmetrical.

Probabilistic Error as a Measure of Control-System Performance

J. ZABORSZKY
MEMBER AIEE

J. W. DIESEL
ASSOCIATE MEMBER AIEE

AUTOMATIC FEEDBACK-control system synthesis occupies a rather que place among the various fields of ineering due to the indirectness of its roach. An antenna will be designed a certain radiation pattern because radiation pattern is called for in intended application. On the other d, a servomechanism which is desd to handle step inputs exclusively be synthesized by the techniques of quist, Nichols, or Bode. This means the system is actually designed for a uency response: to exhibit certain nance characteristics when subjected steady-state sinusoidal inputs of ying frequency. Yet the system will er be subjected to this latter type of it. The justification of this approach hat the experience of the designer vides the basis for expecting that the em will perform properly in its al input environment if it is built a specific frequency response. Such nples of indirectness could be enuated almost endlessly.

n brief, a common characteristic of t design techniques (including here design leading to writing specificas) seems to be that they focus at one at a few restricted aspects of the em operation which by themselves not necessarily significant in the al operating environment of the em. From the study of these aspects designer extrapolates mentally the em performance to be expected in the actual operating environt. Guiding him in this extrapolation his experience, engineering judgment, and some semiquantitative relations, the actual quantitative relationships frequently being too laborious to handle. In addition, the designer is also averaging mentally, on the basis of more or less conscious probability considerations regarding the utilization of the output with various inputs. In other words, while the synthesis is immediately concerned with various special facets of performance under largely hypothetical conditions the real concern is still the performance in the actual environment during the utilization of the output. Thus, looking behind all the indirectness of automatic feedback-control design, it develops that, although largely subconscious, actual system design is based on the one ultimate measure of performance: How much error is there at the times when the system output is utilized?

In this paper an attempt is made to give this measure of performance mathematical formulation which can be the basis of evaluating existing systems or the criterion for designing optimum systems. This measure is believed to be more general and to have more clear-cut physical meaning than existing criteria. In contrast to most standard techniques, which concentrate on isolated phases of performance such as transient or steady state, the proposed measure unites all of these, favoring none, in a single concept.

This paper is intended as the first of a group of papers. Accordingly, the basic philosophies are expounded here, the mathematical formulation and definition of the measure is introduced, and it is demonstrated that these definitions are amenable to mathematical analysis. General solutions rather than working equations are presented.

Proposed Measure of Performance Evaluation

An automatic feedback-control system exists to convert its input into a desired output which may or may not be identical to the input. The difference of desired and actual outputs is the error of the control system. Consequently the performance of the control system is characterized by the amount of the error at such times as the output is utilized. Clearly, this verbal statement of the evaluation of performance contains two elements: 1. the amount or size of the error; and 2. the times when the output is utilized. Both of these elements are subject, in their nature and relative significance, to the environmental conditions of the specific control system. For instance, in some control systems enlarged errors are increasingly, and possibly cumulatively, objectionable. Frequency control of some power systems might be an example. In other cases, such as in an automatic milling machine, errors up to a certain tolerance are not objectionable but all errors exceeding this tolerance are equally and totally objectionable. For most control systems (even for the examples mentioned) the lines are not so clearly drawn and the situation is intermediate between the two extremes. The point made is that the specific environment and function of

Paper 59-199, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., February 1–6, 1959. Manuscript submitted October 23, 1958; made available for printing December 16, 1958.

J. ZABORSZKY is with Washington University and consultant to McDonnell Aircraft Corporation, both St. Louis, Mo., and J. W. DIESEL is with McDonnell Aircraft Corporation, St. Louis, Mo.

each control system will set a specific penalty valuation on errors of a given size. This penalty valuation can be expressed for the given application in the form of a penalty function $F(e)$, a single-valued function of the error e. In a more general sense, the penalty valuation may vary with extraneous parameters such as rate of flow in a temperature control system or with time, $F(e, t, v_1, v_2 \ldots v_r)$, where v_1 through v_r are the parameters. Of course error, e, itself certainly is a function of time, t.

The second element of performance evaluation, the time when the output is utilized, shows equally wide individual variation depending on the environment of the control system. Some control systems, such as a frequency or voltage control on a power system, operate at all times on essentially steady reference inputs and a stationary random noise input, the varying load. All times for utilizing the output are equally probable here. The other extreme would be a control system which makes one adjustment upon activation by a step input followed by one instantaneous measurement after a fixed-time interval. Here the time of utilization is one instant after a fixed-time interval which has probability one while all other time intervals have probability zero. If the measurement itself takes a certain time, all times within this measuring time are equally probable. In a missile, the output is utilized at one single instant, but over a number of launchings the time intervals elapsing from launching to explosion will form a probability distribution centering around the nominal range.

Essentially, then, for any automatic feedback-control system, the times elapsing from activating the system to all times of utilization of its output can be arranged into a probability distribution, $p(t)$, over many repeated activations by identical inputs of the same system or identical systems.

Mathematical Formulation of the Measure of Performance

The measure of performance described in the preceding section can be mathematically formulated as

$$ς = \int_0^{\infty} F(e(t),t,v_1 \ldots v_r)p(t)dt \qquad (1A)$$

where the symbol ς is the end sigma, the form of letter sigma used at the end of Greek words.

For one specific type of input, a "deterministic" input, equation 1(A) clearly gives the statistically correct average value of the penalized error at all times when the output is utilized.

If the environment of the control system includes a number of different types of inputs, possibly an infinite number or ensemble of different types of inputs, an averaging over, this ensemble will also be necessary to obtain the mean penalized error. Then the measure of performance takes the form

$$ς = \int_0^{\infty} \overline{F(e(t),t,v_1 \ldots v_r)p(t)}dt \qquad (1B)$$

where the straight line denotes the averaging over the ensemble, so that equation $1(B)$ describes the average penalized error.

Frequently the probability distribution of utilization $p(t)$ will be independent of the type of the input. Then equation 1(B) reduces to

$$ς = \int_0^{\infty} \overline{F(e(t),t,v_1 \ldots v_r)}p(t)dt \qquad (1C)$$

The value of the proposed measure of performance as defined in equations 1(A)–1(C) is conceptually the average value of penalized error at such times when the output is utilized, and for purposes of identification it will be referred to by either of the names "probabilistic error" or "end sigma error." As a performance measure or criterion equations 1(A)–1(C) will be referred to as the "end sigma criterion."

This measure unites in a single concept the transient and steady states of the system operation as well as the largely neglected intermediate states. None of these operating states is discriminated for or against and because of $p(t)$ each gets the weight due it because of the relative frequency of its presence at such times as the output is utilized.

In addition to uniting the various states of system performance, equation 1 also unites most of the various criteria used to evaluate these states separately. As an example, if $F(e, t, v_1 \ldots v_n) = e^2$, then referring to equation 1(B)

$$ς = \int_0^{\infty} \overline{e^2(t)p(t)}\, dt \qquad (2)$$

becomes the true mean-square value of the error. This end sigma, or probabilistic error, unifies in one concept the mean square criterion used previously for stationary[1,2] and nonstationary[3–6] random input environments and the integral square criterion used for nonstationary random[2] and deterministic[2] environments. These become special cases of equation 2. Final value systems[6] also are special cases of equation 1 by the use of a probability function as given in equation 11. However, equation 2 is not restricted to these special cases. This is general in that it truly averages the square error whatever may be the operating state of the system.

Performance criteria similar in form to equation 1, but basically different in meaning, have been proposed.[7–14] The basic difference is that instead of the probability distribution function, $p(t)$ of equation 1, an arbitrary weighting function is proposed either in a discrete[8–14] or a general form.[13,14] Consequently these criteria define arbitrarily weighted penalty factors without clear physical meaning. Nevertheless, equations 1(A)–1(C) include as special cases these arbitrarily weighted penalty criteria. Assume for instance

$$F(e(t),t,v_1 \ldots v_r) = F(e)W(t)$$

where $W(t)$ is a function weighting, or penalizing, the errors according to the time when they occur. $W(t) = t$ and $W(t) = t^2$ have been proposed and investigated.[8–14] Then, for instance, equation 1(C) reduces to

$$ς = \int_0^{\infty} \overline{F(e)}g(t)dt$$

where

$$g(t) = W(t)p(t)$$

now becomes a modified probability function which, in addition to the relative frequency of occurrence of various utilization times, also includes the designers opinion regarding the relative importance of these utilization times. Equations 1(C) and 4, being identical in form, can be handled by identical techniques which are introduced in subsequent parts of this paper. It is significant that these techniques make it possible to obtain mathematical solutions for the optimum systems under the earlier performance criteria, something which has been done heretofore only by cut and try.

Using equation 1, it is not necessary to define a number of separate performance criteria. Transient and steady conditions, deterministic or random input environments, linear and nonlinear systems are all covered by the same criterion.

Special Cases of Penalty Functions

If the magnitude of the error alone is to be the measure of its undesirability, then

$$F(e) = |e|$$

he penalty function. The widely
l penalty function

$$=e^2 \quad (7)$$

course puts an increasingly high
alty on increasing errors. This
perty does not fit such cases as
matic machine tools or missiles
re all errors up to a certain tolerance
almost equally acceptable and all
rs beyond the tolerance are equally
cceptable. For such control systems

$$=\frac{e^2}{k+e^2} \quad (8)$$

k a constant, would be quite suitable
e this penalty function (Fig. 1)
alizes errors progressively up to a
ain value and then puts an essentially
tant penalty on errors past this
e.

ven more suitable would be the
wing:

$$t))=1-\int_{-e_t}^{e_t} p(e|t)de=p_t(t) \quad (9)$$

h gives the probability of error, e,
g outside the tolerance value, e_t at
t. In equation 9, $p(e|t)$ is the
litional probability distribution in
d t. For the penalty function of
tion 9, equation 1 yields

$$\int_0^\infty p_t(t)p(t)dt=1-$$

$$\int_{-e_t}^{e_t}\int_0^\infty p(e|t)p(t)dedt$$

$$-\int_{-e_t}^{e_t}\int_0^\infty p(e,t)dedt \quad (10)$$

h is simply the over-all probability
e error being outside the tolerance,
st appropriate criterion, which in the
of a missile could readily be extended
e over-all hit probability.

should be emphasized that the
fic $F(e)$ function is determined by
nature of the problem and is to be
blished by the same engineering
ment which establishes countless
r factors of the design. The penalty
tion $F(e)=e^2$ is the most commonly
, and it will be used exclusively in
emainder of the paper.

cial Cases of Probability istribution Functions

definition, $p(t)$ is the probability
ibution of all times when the out-
is being utilized, based on a large
ber of repeated applications of
ts of one given type. The starting
of the time scale $t=0$ for deter-
ng $p(t)$ is the instant of activation

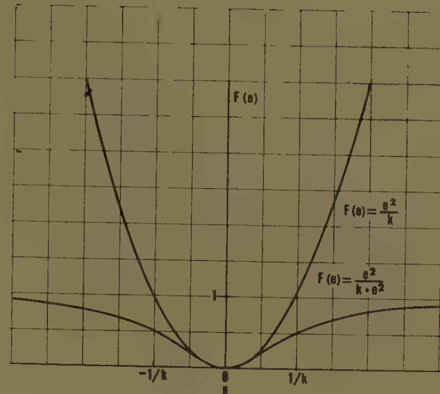


Fig. 1. Two penalty functions

of the control system. For a missile, $t=0$ can signify the instant of firing or the instant of transition to the homing phase. For some other system, $t=0$ may be the instant of turning on the power in the control system or of the arrival of a step input, etc.

The probability distribution function, $p(t)$, can be readily established in some cases. In others, exact determination would require an extensive operations study, but a reasonable estimate can almost always be found quite readily. As a matter of fact, by a slight shift of the sense of the definition, $p(t)$ could be represented as expressing the distribution to which the spread of the possible utilization times can be narrowed down on the basis of the limited information available to the designer. In this sense, $p(t)$ would become, at least partly, representative of the lack of definition in the design parameters rather than a probability distribution of the actual operating times.

As an example, consider a control system which receives step inputs at $t=0$ followed by one instantaneous measurement at $t=T_0$.

Then

$$p(t)=\delta(t-T_0) \quad (11)$$

where $\delta(t)$ denotes the Dirac delta function. If in the same system the measurement extends from T_0 to T_1 (as it would for instance in some wind-tunnel applications)

$$p(t)=\frac{1}{T_1-T_0}[u(t-T_0)-u(t-T_1)] \quad (12)$$

where $u(t)$ denotes the unit-step function.

In writing equations 11 and 12 the assumption was that the times of utilization are the same for every repeated application of the step input. Should these be variable in some way (in the case of a missile the instant of explosion, T_0 in equation 11, would vary from launching to launching depending on the distance of the target at the time of launching) then equations 11 and 12 modify respectively into

$$p(t)=\lim_{K=\infty}\frac{1}{K}\sum_{k=1}^{K}\delta(t-T_{0k})=\overline{\delta\ (t-T^0{}_k)} \quad (13)$$

$$p(t)=\lim_{K=\infty}\frac{1}{K}\sum_{k=1}^{K}\frac{1}{T_{1k}-T_{0k}}\times[u(t-T_{0k})-u(t-T_{1k})]$$

$$=\overline{\frac{1}{T_{1k}-T_{0k}}[u(t-T_{0k})-u(t-T_{0k})]} \quad (14)$$

For a missile where maximum and minimum ranges (T_1 and T_0 respectively) are clearly set, the results of equation 13 would come close to the form of equation 12.

For systems which operate essentially in the steady state, one possible form can be produced from equation 12:

$$p(t)=\lim_{T_1=\infty}\frac{1}{T_1}[u(t)-u(t-T_1)] \quad (15)$$

or from a Rayleigh probability distribution:

$$p(t)\begin{cases}=\lim\limits_{\sigma=0}\dfrac{t}{\sigma^2}\exp\left(-\dfrac{t}{2\sigma^2}\right) & t\geqslant 0\\ =0 & t\leqslant 0\end{cases} \quad (16)$$

With either equation 15 or 16, equation 2 goes over into the customary definition of the rms error as used in the study of systems subjected to stationary random processes. This is especially obvious in connection with equation 15.

Finding the Probabilistic Error for Linear Systems in s Domain

The end-sigma mean-square error is defined by equation 2. Its computation for linear-time-invariant lumped-constant systems can be carried out either in the s domain of the Laplace transforms or in the time domain. The s-domain approach will be discussed first.

The Laplace transform of the error can be expressed as

$$E(s)=I(s)-K(s)R(s)=\frac{M(s)}{N(s)} \quad (17)$$

where $M(s)$ and $N(s)$ are entire functions with $N(s)$ having a higher order singularity than $M(s)$. $I(s)$ and $R(s)$ are, respectively, the Laplace transforms of the desired output and the actual input, and $K(s)$ is the closed-loop transfer function of the system. Now, as shown in the Appendix, equation 2 for a deterministic input $[\overline{e^2(t)}=e^2(t)]$ can be converted by complex convolution into

$$s=\lim_{t=0} \mathcal{L}_2^{-1}\{P(-s)[E(s)*E(s)]\} \qquad (18)$$

where $\mathcal{L}_2^{-1}$ denotes the inverse 2-sided Laplace transform.[15] $P(s)$ is the 1-sided Laplace transform of $p(t)$, and the asterisk denotes complex convolution. Since $p(t)$ is zero for negative time ($t=0$ being the instant of activation, so that any utilization of the output can come only in positive time),

$$P(-s)=\mathcal{L}_2[p(-t)] = \frac{A(s)}{B(s)} = \int_{-\infty}^{\infty} p(t)e^{st}dt \qquad (19)$$

where $A(s)$ and $B(s)$ are entire functions. Further,

$$E(s)*E(s)=\frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} E(s-w)E(w)dw = \sum_i \frac{M(s-p_i)M(p_i)}{N(s-p_i)N'(p_i)} = \frac{J(s)}{L(s)} \qquad (20)$$

where $J(s)$ and $L(s)$ are entire functions, the p_i are the zeros of $N(s)$, all of which are assumed to be single, and

$$N'(s)=\frac{d}{ds}N(s) \qquad (21)$$

By using inverse transformation through contour integration, equation 18 goes over into one of the two alternate forms of equation 22 and equation 24

$$s=-\sum_k \frac{A(q_k)}{B'(q_k)}\frac{J(q_k)}{L(q_k)} \qquad (22)$$

where the q_k are the poles of $P(-s)$, all of which are assumed to be single, and all are located in the right half plane. Also

$$B'(s)=\frac{d}{ds}B(s) \qquad (23)$$

Alternatively,

$$s=\sum_i \frac{A(p_i)}{B(p_i)}\frac{J(p_i)}{L'(p_i)} \qquad (24)$$

where the p_i are the zeros of $L(s)$, all of which are assumed single, and located in the left half plane. Also

$$L'(s)=\frac{d}{ds}L(s) \qquad (25)$$

Two equations (equations 22 and 24) result from the 2-sided inverse transformation since the function is continuous at $t=0$ and, consequently, its limit can be evaluated either from the transform for positive time or from the transform for negative time, in other words, by closing the contour in the s plane on either the right or left.

Equations 22 and 24 represent the general s-plane formulas for evaluating the end-sigma error for time-invariant lumped-constant linear systems and deterministic inputs. The evaluation necessitates the finding of the poles of the error $E(s)$ and of the transform of the probability distribution function $P(-s)$. These are then substituted into the rational function-type expressions of equations 22 or 24. The evaluation of the poles of $E(s)$ usually can be accomplished by root-locus techniques. The substitution into equations 22 or 24 can be aided materially by conventional spirule operations. Inherent in these derivations is the assumption that $E(s)$ has no pole at the origin.

For example, assume an input

$$r(t)=i(t)=1+2t \qquad t>0$$

and probability function

$$p(t)=\frac{10}{3}(e^{-2t}-e^{-5t})$$

while the system closed-loop transfer function is

$$K(s)=\frac{s+1}{s^2+s+1}$$

Then

$$R(s)=I(s)=\frac{s+2}{s^2}$$

$$E(s)=I(s)-K(s)R(s)=\frac{s+2}{s^2+s+1}$$

so that $p_1=-0.5+j\sqrt{0.75}$ $p_2=-0.5-j\sqrt{0.75}$

Also

$$P(-s)=\frac{10}{(s-2)(s-5)}$$

so that $q_1=+2$, $q_2=+5$.

From equation 20

$$\frac{J(s)}{L(s)}=\frac{(s-2p_2)(s-p_1+2)(p_1+2)-(s-2p_1)(s-p_2+2)(p_2+2)}{(s-2p_1)(s-2p_2)(s-p_1-p_2)(p_1-p_2)}$$

and the end-sigma error from equation 22 is $s=1.37$.

This highly simplified example should be interpreted as illustrative of the computation process and not as illustrative of the power of the method attainable when working equations are developed from general solutions given in this paper.

Finding the Value of the Probabilistic Error in the t Domain

If the end-sigma error defined by equation 2 in the t domain is to be evaluated in the time domain, first it will be assumed that the system is time-invariant linear with known impulsive response or weighting function, $k(t)$. Consequently the error can be defined by time-domain convolution, or superposition, as

$$e(t)=i(t)-\int_0^{\infty} k(t_1)r(t-t_1)dt_1 \qquad (26)$$

where $i(t)$ is the desired output, $r(t)$ is the input, and $k(t)=0$ for $t<0$ which is necessary for a realizable $k(t)$.

Let the following function be defined

$$\sigma^2(t,\tau)=\int_0^{\infty} p(t_2+t)e(t_2)e(t_2+\tau)dt_2 \qquad (27)$$

Then

$$s=\sigma^2(0,0) \qquad (28)$$

Substituting equation 26 into equation 27, and rearranging by an exchange of the sequence of integration and considering equation 28, there results

$$s=\psi_{ii}(0,0)-2\int_0^{\infty} dt_1 k(t_1)\psi_{ir}(0,-t_1)+\int_0^{\infty} dt_2 k(t_2)\int_0^{\infty} dt_1 k(t_1)\psi_{rr}(t_0,t_2-t_1) \qquad (29)$$

where the following definitions were made:

$$\psi_{rr}(t,\tau)=\int_0^{\infty} dt_2 \overline{p(t_2+t)r(t_2)r(t_2+\tau)} \qquad (30)$$

$$\psi_{ii}(t,\tau)=\int_0^{\infty} dt_2 \overline{p(t_2+t)i(t_2)i(t_2+\tau)} \qquad (31)$$

$$\psi_{ir}(t,\tau)=\int_0^{\infty} dt_2 \overline{p(t_2+t)i(t_2)r(t_2+\tau)} \qquad (32)$$

where the straight line denotes ensemble averaging for the conditions of equation 1(B).

Equation 29, then, is the time-domain equivalent of equations 22 or 24. The evaluation of the integrals in equation 29 yields the end-sigma error value for a given system. The results are readily generalized to cases of finite memory transfer functions, ($k(t)=0$ if $t>T$), or probability density functions of finite duration, ($p(t)=0$ if $t>T$), by simply changing the upper limit of the integrals to T in equations 1 and 27–32.

Some additional discussion, however, is due equations 30–32. These equations define a generalized form of auto- and cross-correlation functions which unifies in one concept the auto- and cross-correlation functions customarily used for stationary processes[1] and the auto- and cross-translation functions proposed for transient conditions.[2] Clearly the probability function of equation 1

nediately reduces equations 30 through to the definitions of auto- and cross-elation functions. The same prob-ity functions would result in a ified but equivalent definition of the slation functions.[2] An infinitude of special cases result from the various icular probability functions.

aking a time varying Laplace trans-[4] of equations 30 through 32, a eralized power-spectral density cor-onding to the general correlation tions would result.

eferring to equation 30,

$$,s)=\int_{-\infty}^{\infty}\psi_{rr}(t,\tau)e^{s\tau}d\tau$$

$$=\int_0^{\infty}p(t_1+t)\Phi(t_1,s)dt_1 \qquad (33)$$

re

$$s)=\int_{-\infty}^{\infty}\overline{r(t_1)r(t_1+\tau)}e^{-s\tau}d\tau \qquad (34)$$

he apparent power-spectral density the ensemble at time t_1. If the mble is stationary and the ergodic ry applies,

$$s)=\Phi(s) \qquad (35)$$

the generalized power spectral ity,

$$s)=\Phi(s)\int_{-\infty}^{\infty}p(t_1+t)dt_1=\Phi(s) \qquad (36)$$

ces to the conventional power-tral density of the stationary process, of course should. Similar definitions easily obtained from equation 32 a generalized cross-power spectral ity.

ding the Optimum Invariant inear System Under the End-igma Criterion

quations 22, 24, and 29 are useful in uating the end-sigma error for a em of known construction and charac-tics by working respectively in the t domains. If a choice exists between ral potential systems, finding and paring s for each of these immediately pinpoint that particular system h gives optimum error performance, is, minimum true rms error. This ld be an optimum from among a ted choice, established by cut-and-techniques. Now if the choice of ems is made unlimited to the point including all linear systems with tant coefficients, then a direct solu-for the optimum system becomes ible, by finding the optimum $k(t)$ weighting function or its transform, the $K(s)$ transfer function. Since either of these unknown functions is included in the defining integral of s, the quantity to be minimized (equations 18 and 29 respectively), the solution can be obtained by calculus of variations. This will be illustrated here for the time domain, equation 29. The minimum of s is defined by equating its first variation, δs, to zero, that is from equation 29

$$\delta s=2\int_0^{\infty}dt\delta k(t)\times \left[\int_0^{\infty}dt_1k(t_1)\psi_{rr}(t,t-t_1)-\psi_{ir}(0,-t)\right]=0 \qquad (37)$$

which is only possible for arbitrary $\delta k(t)$ if

$$\int_0^{\infty}dt_1k(t_1)\psi_{rr}(t,t-t_1)-\psi_{ir}(0,-t)=0 \quad t>0 \qquad (38)$$

It can be shown that equation 38 actually defines a minimum.

As in the case of equation 29, equation 38 is also easily adapted to finite memory-transfer functions and finite duration probability distribution.

Equation 38 is a linear integral equation of the first kind. $\psi_{rr}(t,t-t_1)$ is its kernel and as may readily be shown from equation 30 this kernel is symmetric. The solution of this integral equation for $k(t)$ is the impulsive response of the optimum system, the system yielding the minimum end-sigma error. The latter then should be found from equation 29.

The literature, both mathematical[16] and engineering,[3] for solving integral equations is extensive. No attempt toward completeness is made in the references.

Conclusions

As a measure of performance, the end sigma or probabilistic error eliminates the indirectness so characteristic of usual design techniques. The end-sigma error summarizes the entire system performance, rather than explaining it in full detail, thus making it a desirable starting point for analysis or design. Such techniques as s-plane transient analysis can then be used to spotlight special limited aspects of the performance.

In addition to introducing this measure of performance it has been demonstrated that it can be evaluated in either the s or t domain for time-invariant linear systems. An integral equation of a common type has been introduced to define the impulsive response of the optimum-invariant linear system under the end-sigma criterion.

Appendix

Equation 2 can be rephrased as

$$s=\lim_{t=0}\int_0^{\infty}p(\tau-t)e^2(\tau)d\tau \qquad (39)$$

Or introducing 2-sided Laplace transformation[16,17] $\mathcal{L}_2[\ ]$ and inverse 2-sided Laplace transformation $\mathcal{L}_2^{-1}[\ ]$

$$s=\lim_{t=0}\mathcal{L}_2^{-1}\ \mathcal{L}_2\left[\int_0^{\infty}p(\tau-t)e^2(\tau)d\tau\right]$$
$$=\lim_{t=0}\mathcal{L}_2^{-1}\left[\int_{-\infty}^{\infty}e^{-st}dt\times \int_0^{\infty}p(\tau-t)e^2(\tau)d\tau\right]$$
$$=\lim_{t=0}\mathcal{L}_2^{-1}\left[\int_0^{\infty}e^2(\tau)e^{-s\tau}d\tau\times \int_{-\infty}^{\infty}p(\tau-t)e^{s(\tau-t)}d(t-\tau)\right] \qquad (40)$$

Further, from complex convolution theory

$$\mathcal{L}_2[e^2(\tau)]\int_0^{\infty}e^2(\tau)e^{-s\tau}d\tau=E(s)*E(s) \qquad (41)$$

Equations 40, 41, and 19 can be combined into equation 18.

References

1. Extrapolation, Interpolation, and Smoothing of Stationary Time Series (book), N. Wiener. John Wiley & Sons, Inc., New York, N. Y., 1949.

2. Analytical Design of Linear Feedback Controls (book), G. C. Newton, Jr., L. A. Gould, J. F. Kaiser. John Wiley & Sons, Inc., 1957.

3. Random Processes in Automatic Control (book), J. H. Laning, Jr., R. H. Battin. McGraw-Hill Book Company, Inc., New York, N. Y., 1956.

4. An Extension of Wieners Theory of Prediction, L. A. Zadeh, J. R. Ragazzini. *Journal of Applied Physics*, New York, N. Y., vol. 21, 1950, p. 645.

5. An Optimization Theory for Time-Varying Linear Systems with Non-Stationary Statistical Inputs, R. C. Booton, Jr. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 40, Aug. 1952, p. 977.

6. Final Value Controller Synthesis, M. V. Mathews, C. W. Steeg. *Transactions*, Professional Group on Automatic Control, Institute of Radio Engineers, New York, N. Y., vol. PGAC-2, Feb. 1957, p. 6.

7. A Note on Control Area, T. M. Stout. *Journal of Applied Physics*, New York, N. Y., vol. 21, 1950, p. 1129.

8. Some Design Criteria for Automatic Controls, Paul T. Nims. *AIEE Transactions*, vol. 70, pt. 1, 1951, pp. 606–11.

9. The Synthesis of "Optimum" Transient Response: Criteria and Standard Forms, D. Graham, R. C. Lathrop. *Ibid.*, vol. 72, pt. II, *Applications and Industry*, Nov. 1953, pp. 273–88.

10. The Transient Performance of Servomechanisms with Derivative and Integral Control, Richard C. Lathrop, Dunstan Graham. *Ibid.*, vol. 73, pt. II, Mar. 1954, pp. 10–17.

11. THE INFLUENCE OF TIME SCALE AND GAIN ON CRITERIA FOR SERVOMECHANISM PERFORMANCE, Dunstan Graham, Richard C. Lathrop. *Ibid.*, July, pp. 153–58.

12. AN INTEGRAL CRITERION FOR OPTIMIZING DUPLICATOR SYSTEMS ON THE BASIS OF TRANSIENT RESPONSE, Jack Crow. *Doctoral Dissertation*, Washington University, St. Louis, Mo., 1956.

13. THE SELECTION AND USE OF SERVO PERFORMANCE CRITERIA, W. C. Schultz, V. C. Rideout. *AIEE Transactions*, vol. 76, pt. II, 1957 (Jan. 1958 section), pp. 383–88.

14. A GENERAL CRITERION FOR SERVO PERFORMANCE, W. C. Schultz, V. C. Rideout. *Proceedings*, National Electronics Conference, Chicago, Ill., vol. 13, 1957, pp. 549–60.

15. THEORIE AND ANWENDUNG DER LAPLAC TRANSFORMATION (book), G. Doetsch. Dove Publications, Inc., New York, N. Y., 1943.

16. LINEAR INTEGRAL EQUATIONS (book), W. V Lovitt. Dover Publications, Inc., 1950.

17. AUTOMATIC FEEDBACK CONTROL SYSTE SYNTHESIS (book), J. G. Truxal. McGraw-Hi Book Company, Inc., 1955.

Modern Systems of Traffic Control as Applied to the Seaboard Air Line Railroad Company

J. R. DEPRIEST
NONMEMBER AIEE

ON December 6, 1941, the Seaboard Air Line Railroad placed in service the first segment of a major installation of a system of traffic control on its line. The segment was only 33 miles in length but it made history. It was a portion of a 65-mile segment of railroad to be electrically controlled from a remote location 157 miles from the most distant point. At that time this line became the longest control line for this type of service in the world. The system proved to be very efficient and economical, and expansion of the system was made as rapidly as material and labor could be obtained. As of today, trains operate the entire distance between Richmond and Miami–Tampa by signal indication controlled by modern systems of traffic control.

Prior to December 6, 1941, the Seaboard line consisted of a single-track railroad with more than 4,000 miles of track extending from Richmond to Miami and Tampa. A main line also extended from Hamlet, N. C., toward Atlanta and Birmingham together with other extensions operating in the states of Virginia, North Carolina, South Carolina, Georgia, Alabama, and Florida. At that time the Seaboard had approximately 550 miles of automatic signals mainly extending between Richmond, Va., and Monroe, N. C.; between Savannah, Ga., and Baldwin, Fla.; and between Starke and Coleman, Fla.; see Fig. 1. With the systems then in service it was necessary for telegraph operators or telephone operators to be placed at strategic points along the railroad in order to receive train orders issued by a train dispatcher. The operator then physically delivered these orders to train crews in order that opposing trains could meet at strategically placed sidings at the proper time. Since it takes considerable time to issue and check train orders, it can be seen that the system becomes difficult and expensive to operate as the number of trains increases. This is so because the total number of meets in a fixed distance increases directly in proportion to the square of the number of trains. Perhaps, this point should be further discussed so that the reader may become familiar with the problems encountered in designing a traffic-control system for operating a single-track railroad.

It is evident that for a double-track railroad it is possible to operate trains of equal speeds at equal time intervals without interference from either following or opposing trains, since trains moving in opposite direction operate on adjacent tracks. Fig. 2(A) is a time-distance chart showing the position of southward (*1S*) and northward (*1N*) trains. Train *1S* leaves station *A* proceeding toward station *C* at a constant rate of speed. The progress of this train is represented by a diagonal line with a negative slope. At the same time train *1N*, represented by a diagonal line with a positive slope, operating in the opposite direction, leaves station *C* proceeding toward station *A* at the same speed. Since each train is operating at the same constant speed they will arrive at their destinations *C* and *A*, respectively, after a time interval *t*, having traversed the distance *D* between stations. It will be noted from Fig. 2(A) that the two diagonal lines, representing the progress of train *1S* and *1N*, intersec or cross one another half way betwee stations *A* and *C*, after one half of th time interval *t* has elapsed. Since trai *1S* leaving station *A* will meet only on train between stations *A* and *C* (and on train at *C* if the railroad continues for greater length) the railroad could b single track with the exception of small distance near the center (siding *B* and at stations *A* and *C* if means were provided whereby one of the trains could enter, pass the opposing train, and then leave the siding. It should be pointed out that if train *1S* continued beyon station *C* another train would be met a station *C* if opposing trains were di patched at equal intervals.

Similarly Fig. 2(B) illustrates th conditions encountered if the number trains operated in time interval *t* wer doubled. It can be seen from Fig. 2(B that the total number of meets occurrin within a time interval *t* would increas from two to eight. However, fou sidings would be required as compare with two sidings previously. If trains in each direction were operate within time interval *t*, then by referrin to Fig. 2(C) it can be seen that a total 18 meets would occur within time interva *t* and 6 sidings would be needed.

It is interesting to note that where on train in each direction is operating pe time interval *t*, siding *B* is used one tim as shown by the heavy dot locate vertically under siding *B*. Where tw trains in each direction are operatin per time interval *t*, siding *B* is use twice. And where three trains ar operating in each direction per tim interval *t*, three trains use siding *B*.

If *t* is made equal to a time interval o one day and Table I is prepared fro the foregoing information, a formul can be written from the first three cond tions to determine the total number

Paper 59-251, recommended by the AIEE Lan Transportation Committee and approved by th AIEE Technical Operations Department f presentation at the AIEE Winter General Meetin New York, N. Y., February 1–6, 1959. Manu script submitted November 5, 1958; made availab for printing December 16, 1958.

J. R. DEPRIEST is with the Seaboard Air Lin Railroad Company, Richmond, Va.

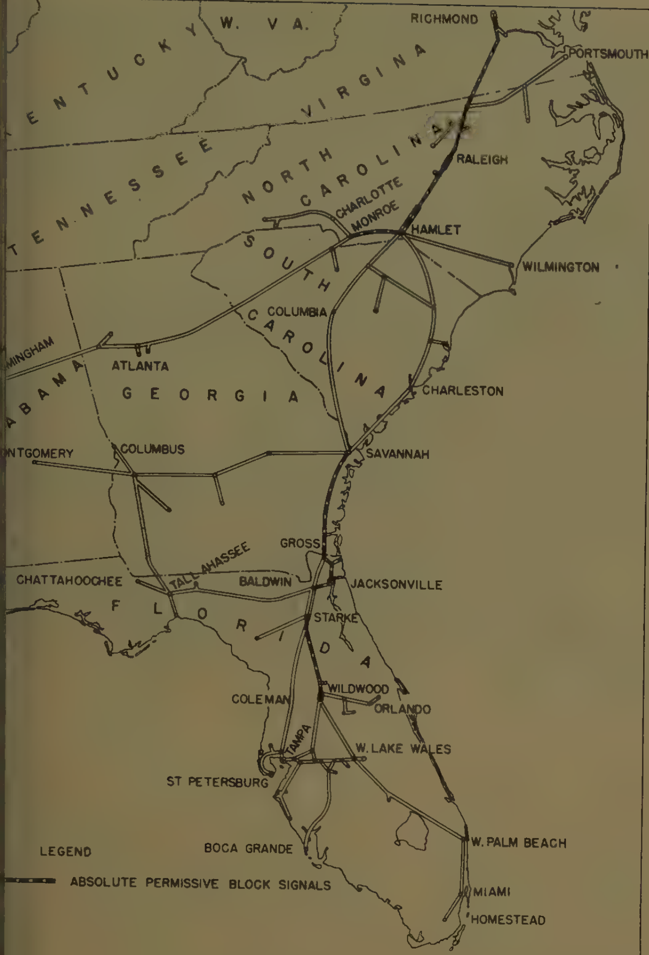


Fig. 1 (left). Type of signals on Seaboard prior to Dec. 6, 1941

Table I

T	t	M	S
1	1	2	2
2	1	8	4
3	1	18	6
4	1	32	8
5	1	50	10
10	1	200	20
20	1	800	40
24	1	1,158	48
48	1	4,608	96
72	1	10,368	144
96	1	18,432	192

ts occurring within the area the
n operates in one day or for any
interval selected. Similarly, for-
as can be derived showing the
ts per train per day which equals the
ber of sidings needed within the
. These formulas are as follows:

$2(Tt)^2$

I/Tt

Tt

where

t = time interval a train operates in traveling distance D

T = number of trains operated per direction per time interval t

M = total number of meets occurring within time interval t or in distance D

D = distance train travels in time interval t

S = number of sidings needed within distance D or time interval t

For a practical example, assume t is equal to one day and one train is operated each half-hour, then 48 trains will be operated in each direction per day. From Table I, under S, 96 sidings will be needed. If the trains average 40 mph, a train will travel 960 miles in one day, which is distance D. Therefore, if the 960 miles is divided by the 96 sidings it will be found that sidings must be spaced 10 miles apart. It should be pointed out that with sidings spaced 10 miles apart and trains averaging 40 mph, if one of the trains was delayed and could not make the scheduled progress, it would be necessary to require the delayed train to enter the siding one station before its scheduled point to avoid delay to the opposing train. This train would then be further delayed for a time interval not exceeding the sum of its running time to the next station and the running time of the opposing train from the next station to the siding where the meet will occur. Where the speeds of each train are equal, then the maximum additional delay that could occur would be two times the running time between sidings, or in this case 18 minutes. To reduce this type of delay more sidings are usually provided than required to handle the maximum number of trains as shown by Table I.

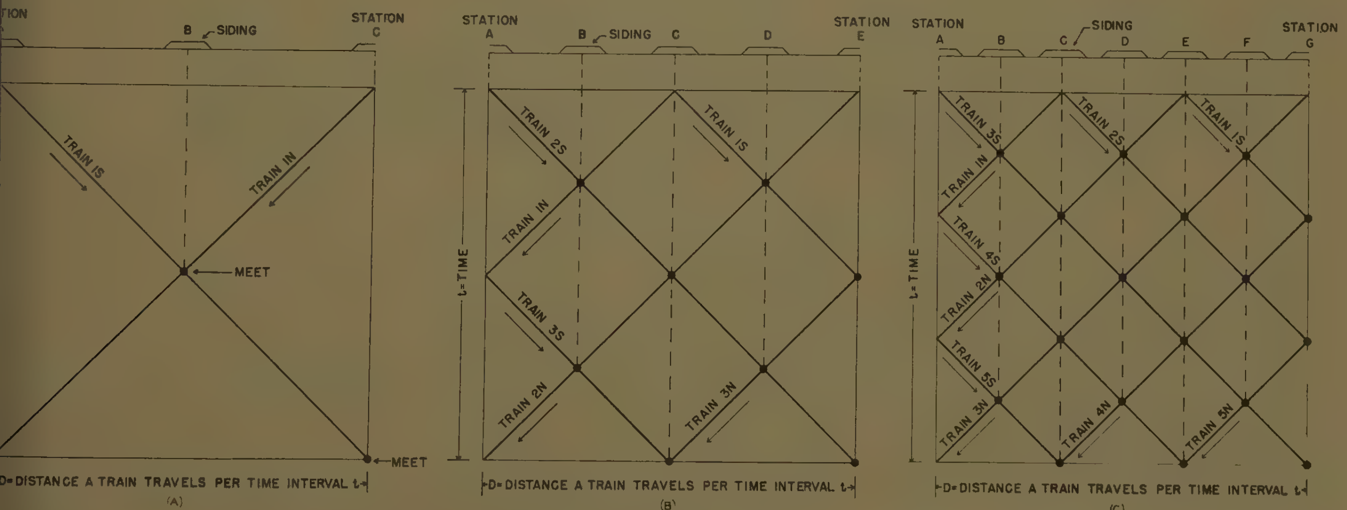


Fig. 2. Time-distance chart

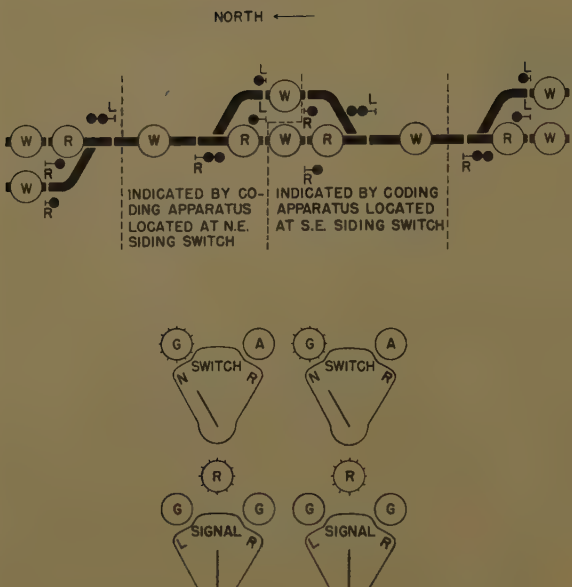


Fig. 3. Track diagram with associated indication lights and control switches

A traffic-control system remotely controlled must provide electric power to operate and control a railroad switch located at the entrance or exit of a siding, and extend information to enginemen of trains by means of wayside signals appropriately located. It is necessary for the dispatcher to control the following functions for each siding switch:

1. Position switch normal.
2. Position switch reverse.
3. Position northward signals to either stop or clear.
4. Position southward signals to either stop or clear.
5. Position automatic reclearing feature for following trains.
6. Operate maintainer's call light.

The system must indicate certain information from the field location, where the power-operated switch is located, to the dispatcher located at control machine at a remote location. The following information is required from each power-operated siding switch:

1. Occupancy or nonoccupancy by train located on track north of switch.
2. Occupancy or nonoccupancy by train located on switch.
3. Occupancy or nonoccupancy of train located south of switch either on main line or siding. (Usually a typical procedure is to assign siding occupancy indications to apparatus located at north switch of siding and occupancy of track between switches to apparatus located at south switch.)
4. Indication to show when switch is set for main-track traffic or in transit position.
5. Indication to show when switch is set for side-track traffic or in transit position.
6. Indication of northward signal aspect (clear or stop).
7. Indication of southward signal aspect (clear or stop).
8. Power off indication (on or off).

Fig. 3 illustrates a track diagram with associated indication lights and control switches.

From the foregoing it will be noted that 6 bits of information are needed for control purposes and 8 bits of information are needed for indicating purposes at each siding switch located in traffic-control territory. The Union Switch *506-A* system, described by Mr. Baughman[1] in another paper, is arranged to control and indicate by code 7 bits of information for 35 stations over one pair of wires. From the foregoing requirements it can be seen that, with the exception of indications, this system will be adequate. Since power is off a small percentage of time, it is practical to indicate both a normal and reverse switch condition simultaneously, and since this condition is physically impossible a means is established to show that the a-c power is off. Since this system uses a share-time basis for controls and indications, it is usually the practice not to control more than 30 switches from one line, or 15 sidings. Experience has proved that this number of sidings can be controlled without unduly delaying indication or control codes for all stations on a system used by a busy railroad. If more than 15 sidings are needed, the territory is divided in two or more sections and a system is applied to each section. When this is done, a pair of wires extending from the most distant siding switch to the control point is necessary for each section of line to which a system is applied. If the d-c coded impulses at the end of each section are used t

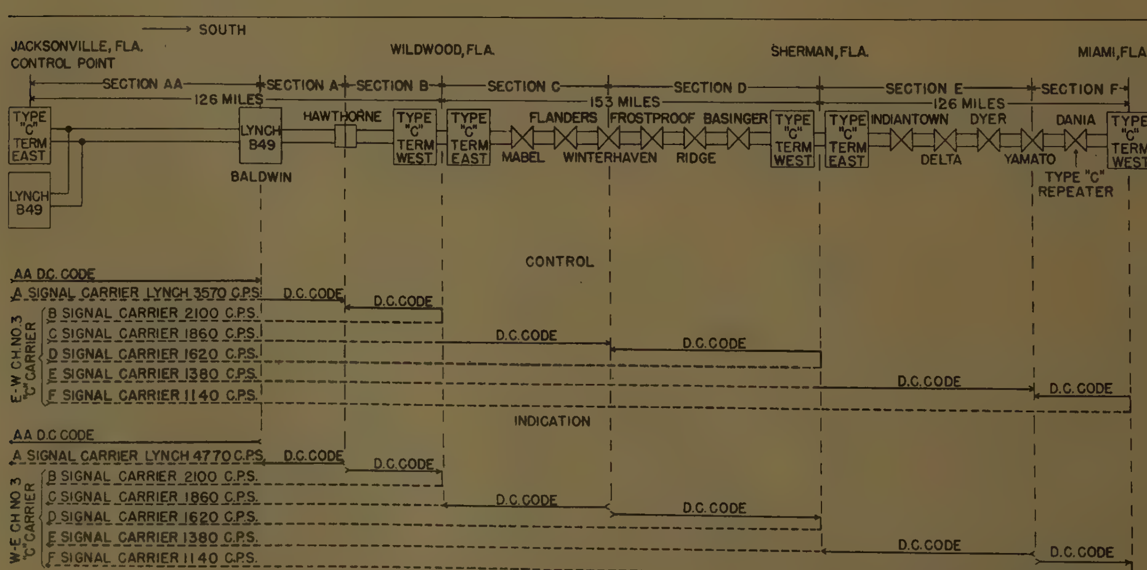


Fig. 4. Diagram ... carrier sections ...

Fig. 5. Traffic control machines located in Jacksonville

ılate a voice-frequency telegraph- carrier, only one pair of wires be needed throughout the entire ad for this purpose. For control , similar carrier apparatus is used he modulation takes place at the ol machine and the carrier is dulated at an end of the remote n in the field. Fig. 4 illustrates the ol arrangement used by the Sea- Air Line Railroad between Jack- lle and Miami, Fla.

will be noted in the foregoing ar- ment that a Western Electric *C* r which provides three trunk tele- circuits was applied to the single of wires of the code line and cable ding between Jacksonville and Wild- , Wildwood and Sherman, and nan to Miami. One channel of arrier was used to apply the voice- ency signal-control carriers. The line or system called *AA* extends een Jacksonville and Baldwin. The narked *A* extends between Baldwin Hawthorne with carrier controls ding this line into Jacksonville its north end at Baldwin. The line designated *B* extends between horne and Wildwood. Since the rier terminal is located at Wild- the south end of this line is ex- d to Jacksonville by carrier. The esignated *C* extends between Wild- and Winter Haven. This line is led from the north end by carrier Jacksonville. Line *D* extends en Winter Haven and Sherman and carrier *C* terminal is located at Sherman, it is extended from the south end to Jacksonville by carrier. The next line designated *E* extends between Sherman and Yamato and the north end is extended by carrier from Sherman to Jacksonville. The last line designated *F* extends between Yamato and Miami and since the *C* carrier terminal is located at Miami the *F* is extended from its south end to Jacksonville. The distance from Jacksonville to Miami is approximately 405 miles. The controls for this entire territory are placed on three machines, one controlling the territory between Jacksonville and Baldwin, another between Baldwin and Wildwood, and the other controlling the territory between Wildwood and Miami. Fig. 5 shows the traffic control machines located in Jacksonville.

Up to this point the communications involved have been discussed, but not the actual wayside signal systems available for traffic control. On the Seaboard three distinct types of systems have been installed. The first system is referred to as the full-centralized traffic control system which permits trains to follow one another without waiting until the preceding train has reached the next siding. This system is equipped with power-operated switches and, where required, the sidings are usually more closely spaced than in the other systems described later. The next system has been designated as the modified-centralized traffic control system. This system provides for power-operated switches, but will not permit trains to follow until the preceding train has reached and passed the entrance switch to the next siding. The system eliminates the many wayside signals needed when following movements are permitted as in the full-centralized traffic control. The third type of centralized traffic control is designated by the Seaboard as remote-controlled automatic signals system. This system basically is arranged for train operation by signal indication in the same manner as the modified-centralized traffic-control system, except that following trains cannot follow another train until the first train has reached and passed the north end of the siding ahead. It can be seen that this system has eliminated much apparatus at one end of each siding. The siding switches are usually hand operated. However, the system is so designed that power-operated switches can be added when needed at a minimum of expense and the system can be expanded into either modified-centralized traffic control or full-centralized traffic control by adding the necessary apparatus for such system. In addition to the traffic control systems the Seaboard has short stretches of double track where the conventional automatic signals are provided and where trains operate by signal indication with current of traffic. Fig. 6 shows the various types of centralized traffic control used on the Seaboard as of October 1958. In addition the map shows the absolute-permissive block signal system between Hamlet and Atlanta. This system is a conventional automatic-permissive block signal system and is used with train orders for the movement of trains. It will be noted from this map that the Seaboard operates by signal indication for the entire distance of more than 1,040 miles between Richmond, Miami, and Tampa.

Three control machines are located at Raleigh, N. C., which control the entire territory between Richmond, Va., and Hamlet, N. C. Three control machines are located at Savannah, Ga., which control the territory between Hamlet via Columbia to Savannah; Hamlet via Charleston to Savannah; Savannah–Gross–Jacksonville and Gross–Baldwin, respectively. At Jacksonville there are three traffic control machines, which control the entire territory between Jacksonville and Miami. One machine controls the Jacksonville-Baldwin section, one the section between Baldwin and Wildwood and the other the section between Wildwood and Miami. The section between Coleman and Tampa, Florida, is controlled from a

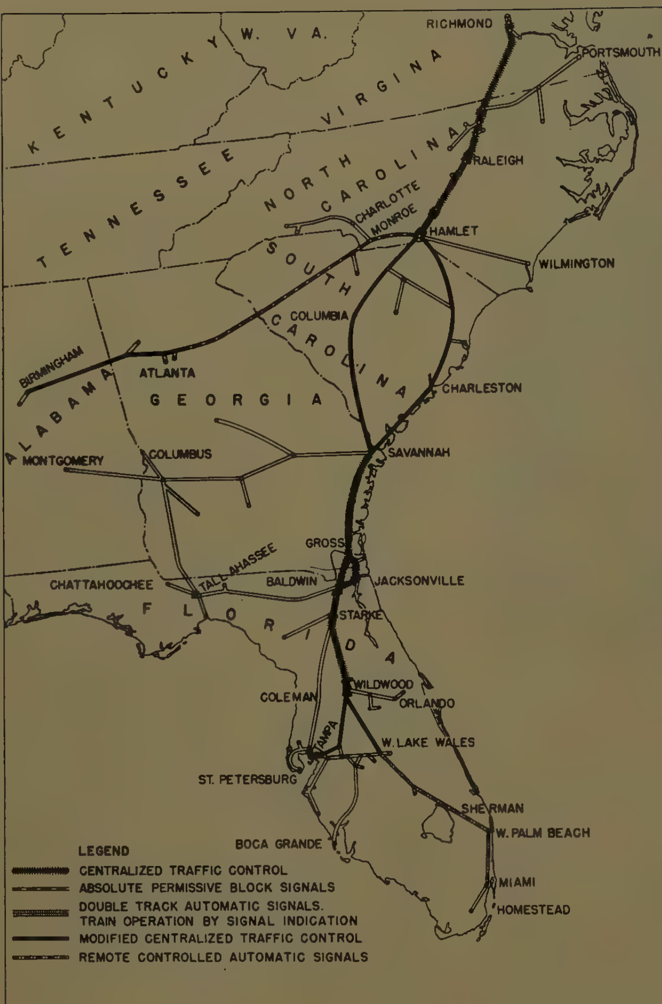


Fig. 6. Various types of centralized traffic control used on Seaboard as of Oct. 1958

machine located in Tampa. Incidentally, at the present time the control line extending between Jacksonville and Miami is the longest centralized traffic-control line in the world to extend in one direction from the control point. Also the system between Richmond and Miami is the longest continuous system in the world on which trains are operated by signal indication.

After the war, in 1946, when materials were again available the Seaboard undertook the installation of traffic control on many of its main lines where signals were not in service. The territories to be controlled were long and a reliable code line was needed. Methods were developed to economically install an underground cable consisting of two no. 10 American Wire Gauge solid copper wires, each circumscribed by 3/64-inch polyethylene insulation. The two polyethylene insulated conductors were twisted with a 4/64-inch Flamenol jacket covering the twisted pair. Details of this installation were given in another paper.[2] Part of this cable has been in service for 11 years, and its performance has been satisfactory with the exception of that cable installed in certain portions of Florida where the ground was infested with pocket gophers. In this territory the pocket gophers, which normally feed on roots, attacked and ate the cable damaging the insulation and eventually causing grounds. The resultant electrolysis eroded the conductors. Experience has proved that the same cable with a 10-mil bronze tape wrapped spirally around the cable affords an armor which protects the cable from pocket gophers. Evidence has been found that pocket gophers gnawed the armored cable but were unable to penetrate the armor. Of all the armored cable installed only one failure has occurred and that as a result of a nail having penetrated the armor and insulation, evidently while on the reel before being installed. It is interesting to note that no lightning failures on the armored cable have occurred but on the nonarmored cable an occasional failure does occur due to lightning, even though a test sample of the cable after 6 years of service was examined and found to have a breakdown voltage in excess of 40,000 volts to ground.

It should be pointed out that with systems of centralized traffic control the wayside signal apparatus used is of high quality. The electric circuits are designed on a "fail safe" basis and, even though the communication facilities controlling such wayside signs may fail to operate properly, the wayside signals are arranged to check that all power-operated switches are mechanically locked, all hand-operated switches are in the proper position, rails are intact, the track ahead is clear for either following or opposing trains, and that the signals governing opposing movements have been at stop for the prescribed predetermined time. In the traffic-control systems installed recently on the Seaboard, in the event of communication failure the train automatically will clear signal in advance if the route ahead is clear. This feature is very desirable and eliminates a large number of delays if a communication failure should occur and it meets the company's first and last requisite which is safety.

References

1. Traffic Control for Railroads, G. W. Baughman. *AIEE Transactions*, vol. 78, pt. II (*Applications and Industry*), July 1959, pp. 173-7

2. Carrier Operation in Buried Cable, J. J. DePriest. *Proceedings*, Association of American Railroads, New York, N. Y., Communication Sections, Sept. 1948, pp. 374–88.

Traffic Control for Railroads

G. W. BAUGHMAN
FELLOW AIEE

IE HISTORY of the United States hows in some degree a direct relation- between the country's progress and ouilding of the railroads. One of the oads' most outstanding achieve- ts was the phenomenal manner in h they handled more than 90% of all onnel and materials that were moved nilitary purposes during World War One objective of this paper is to show the centrallized traffic-control system ed the railroads achieve this out- ding performance.

though not generally realized, the oads have been among the leaders in ting and developing technological nces. This paper will mention ly some of the advances which are rical or electronic in nature. The lopment of the traffic-control system, h has meant a great deal to the rail- s, will then be described.

e railroads adopted the telegraph after its invention and developed ods of train operation based upon raphic orders. In a similar manner elephone was adopted. At present ailroads make extensive use of the ern techniques in communications.

e first use of electronic principles commercial scale, outside the fields e telephone and the telegraph, is ted to the automatic cab signal. lopment work on this system started 16, and the first commercial installa- was made in 1923. It is this system causes a signal to be relayed to cab indicating the safety of traffic tions in advance. An important ap- tion of this system will be referred ter.

e transition from steam to the diesel- ic locomotive has taken place at a omenal rate on the railroads in this try. The change is approaching letion with approximately 27,000 -electric locomotives in contrast to 3,000 steam locomotives now in ce.

e railroads are installing large classification yards for the handling of freight cars. These installations include methods for switching each car to its proper destination track in response to previously stored information. This storage may be on teletype tape or other suitable memory system. Radar devices are used for speed measurement in servo loops for the control of the speeds of the cars. These installations are among the nearest approaches to full automation in service today.

Description and Discussion

The preceding indicated some of the electrical developments that have been adopted and developed by the railroads. The paper shall now discuss centralized traffic control (CTC). This is actually a system of train operation where the train crew receives no written orders from the dispatcher. The train is operated in accordance with the signal indications. Train orders may be considered as issued by the dispatcher because he is given remote control of the signals and switches in his territory. A basic concept of this system is that it is divded into two main parts. The first is a code communication system so the dispatcher may convey his desires to the signals and switches. Also this communication system transmits information to the dispatcher giving the positions of signals and switches as well as the location of all trains. The second part consists of a complete wayside signal system with the main signals normally in the "stop" position. The over-all safety resides to a major extent in this part of the system. There are two conditions that must be satisfied before a signal can clear to authorize the passage of a train. The dispatcher must send a code to the particular signal, in effect requesting it to clear. This request is received and decoded automatically and thus the first condition is satisfied. The second condition involves the automatic signal equipment which functions continuously in response to traffic conditions. Before the particular signal can clear, the automatic signaling equipment must check that it is safe for the signal to clear. In a similar manner the dispatcher has control of the track switches in his territory. But before any track switch changes its position, in response to a code from the dispatcher, the automatic signaling equipment must check that it is safe for the switch to move.

Fig. 1 shows a machine at Savannah, Ga., for the control of 250 miles of single-track railroad. A diagram on the machine shows the track layout over which the dispatcher has control. Small lights on the diagram show the positions of all trains and these positions are recorded on a chart usually called an automatic train graph. Also the positions of all switches and signals are indicated by lights on the panel.

The dispatcher sends codes to the various locations in his control of the traffic. This is accomplished merely by the operation of small levers and push buttons. Control codes are generated automatically in response to the positions of the levers and push buttons. Indica-

Fig. 1. A 25-foot control machine in service at Savannah, Ga.

59-252, recommended by the AIEE Land portation Committee and approved by the Technical Operations Department for pres- n at the AIEE Winter General Meeting, ork, N. Y., February 1–6, 1959. Manuscript ted November 5, 1958; made available foi g December 4, 1958.

BAUGHMAN is with the Union Switch & Division of Westinghouse Air Brake Com- Swissvale, Pa.

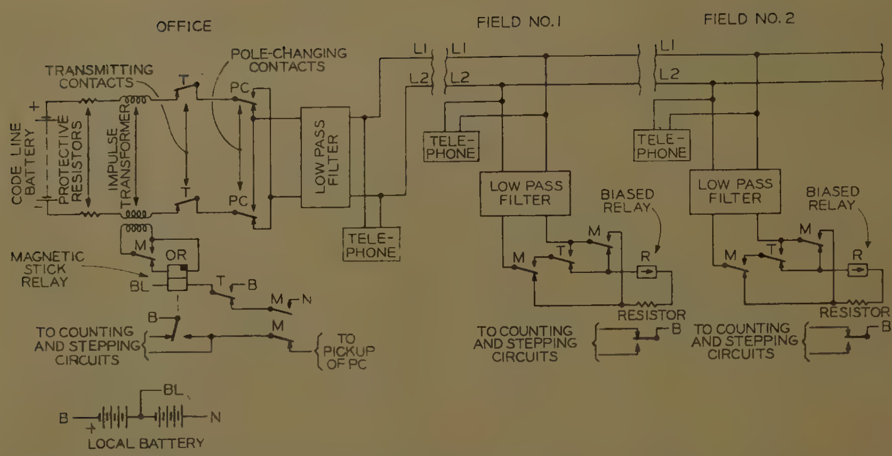


Fig. 2. A multiple CTC line circuit with low-pass filter to permit a physical voice circuit on the same pair of line wires

tion codes from the way stations to the control office are generated in response to the positions of the functions in the field.

Two line wires are used for the transmission of these codes. The use of a single wire with ground return received consideration but was discarded promptly because of the requirement of co-ordination with communications lines. Another reason against the single wire with ground return is the interference that would be experienced during magnetic storms.

The usual installation on a single-track railroad has the field functions arranged in groups of about 6 to 10 functions per station. The stations are spaced along the line at intervals in the order of 5 miles. In usual communications language this might be called a party line. In this application there are frequently 15 to 25 parties on a line.

Fig. 2 shows a simplified line circuit. In the event of simultaneous starts at office and field, the office is given preference by making the first pulse of each control code long as compared with the first pulse of each indication code. The method for establishing priority among the field stations is dependent upon the specific assignment of station codes and the geographical sequence of the stations with respect to the office.

The codes consist of pulses of long and short intervals. The first portion of a code is assigned to station selection and the last portion to the functions of that station.

The main novelty of Fig. 2 with respect to previous line circuits is that the way stations are connected to the line in multiple. The two line wires are continuous throughout the installation whether the line is idle or handling codes. Way stations transmit indication codes by alternately shunting and removing the shunt from the line. This action produces changes in current from the line battery at the office. These changes in current through the indication transformer induce voltages in its relay winding for operation of the office line relay.

Because of the over-all reliability required of this system special attention is given to the strength and ruggedness of the line wires. This being true it is logical to consider the use of these same line wires for other functions. This principle develops in a cumulative manner. The stronger and more rugged the line wires, the greater is the desirability of placing additional functions on that same pair of wires. Also the greater the number of functions being handled over the one pair of wires, the greater is the attention that is warranted in providing reliability and good maintenance.

A major step in permitting additional functions to be handled on the same pair of line wires is the installation of low-pass filters between each way station equipment and the line. The low-pass filter for this service has low resistance to the flow of direct current and high impedance to frequencies in the voice and ultr sonic ranges. The filter is designed t minimize the electrical noise which is fe to the line because of the coding contact To maintain ruggedness and reliabilit of the line circuit, special attention given to this filter so that it will withstan the severe surges that develop on lines o this type during storms. This special a tention to the filter does much to protec other apparatus at the station fro damage by lightning surges. With thes low-pass filters in place, as shown in Fi 2, the pair of wires for the CTC line serve as an excellent voice communications ci cuit. It is the usual practice to have telephone at each way station so tha maintenance men and train crews ma communicate directly with the dispatche

The system as described up to th point was available in 1940, largely b cause of the intelligent use of funds an personnel for intensive development wor during the depression years of the Thi ties. During the early days of World Wa II, the traffic congestion on the singl track railroads was so great that urge studies were made to determine the be way to relieve the congestion prompt and consistent with the economic use strategic materials. The traffic contr system was determined as the most feas ble solution. Emergency measures an much overtime work by the manufa turers and the installation crews provide relief at the most critical locations.

The system operated satisfactorily practice, but soon after there was a d mand for further expansion beyond th approximately 75-mile limit of the syste as it existed. The experience indicate that one dispatcher could handle muc more territory and that additional adva tages would be realized with larger co centration of controls.

The solution to permit large install

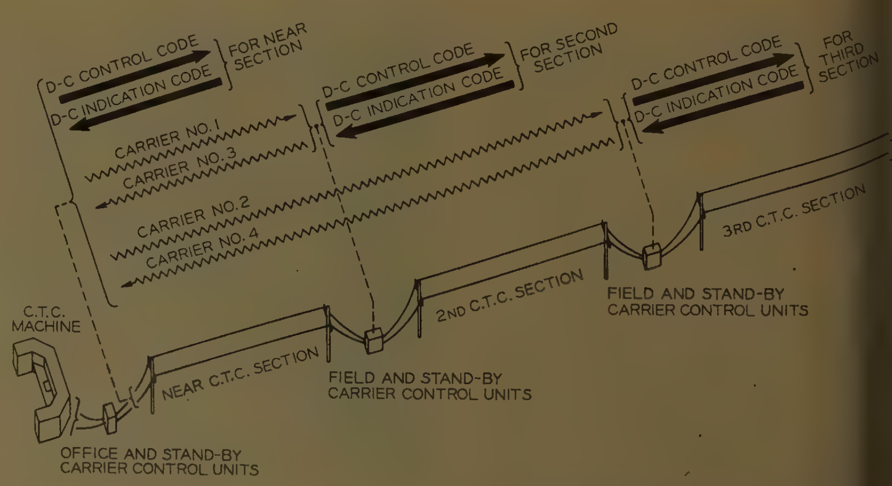
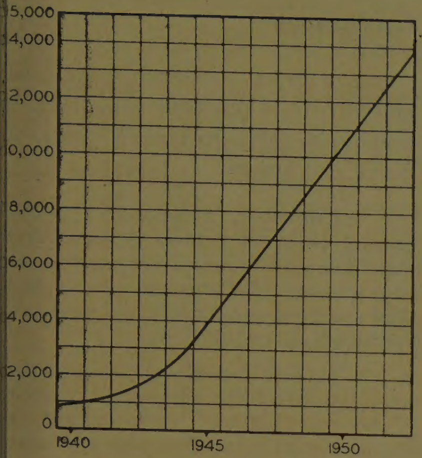


Fig. 3. How carrier frequencies permit the expansion of the CTC control systems



4. **Total road miles of traffic control versus years**

ns, hundreds of miles in length, to be
rated from one control office gave full
sideration to the desirability of retain-
as standard the system and equipment
t had been giving excellent service.
er a great deal of study of various ap-
aches to a solution, the one that
ved to be most acceptable considered
system as it then existed as a unit of a
ger system.

ig. 3 shows three such units placed
to end and controlled from one office.
e near unit operates in the conventional
nner, previously described. The sec-
unit is arranged in exactly the same
nner as if there were a control point
ted at the far end of the first unit.
carrier frequency transmitted in
h direction over the first unit permits
second unit to be controlled over the
wires of the first. In a similar man-
a pair of carrier frequencies over the
wires of the first and second units
nits the handling of a third unit.
filters which were provided to per-
a telephone communication system
e operated on the one pair of wires
serve to permit the transmission of
carrier frequencies. This use of car-
frequencies removed the previous
tations as to size of the installations.
roads may now handle any desired
of installation from one control point.
outstanding example is the Savannah
e on the Seaboard Air Line railroad
which over 600 miles of track are
rolled.

o retain the telephone circuit on the
e pair of line wires it was necessary to
t the frequencies of the carriers so
they would not be audible to the ex-
of causing interference or annoy-
to telephone conversations. Tests
experience indicated 11 kc to be the
mum frequency that was practical.
interesting to note that the first carrier units for this type of service were built on a temporary basis to serve until more substantial units could be made. These units were installed on the Southern Pacific in 1942 and are reported to be continuing in satisfactory service at the present time. To minimize the delay to trains that might be caused by the failure of electronic components the carrier units were installed in duplicate.

Starting with the first carrier units there was an attempt to simplify the manufacturing and inventory problems by making as large a portion of the unit as possible common to all frequencies. The frequency determining units were made separately as plug-in devices.

Special attention should be given to the way that the use of carrier frequencies expanded the size and scope of CTC. All of the main items retained the same use as in the previous system. The addition of carrier frequency and coordinating equipment broke through the previous barriers of the size of installation and the distance of transmission. This indicates why it was possible to meet the sudden demands for larger installations. An idea of the sudden change in demand is indicated by noting the change in slope of the curve in Fig. 4 during the years 1942 to 1943.

An important application of the equipment and system just described, outside the railroad field, was in the extensive remote control and indication system for some of the vital experiments in the Manhattan Project.

An interesting demonstration was performed jointly by the Pennsylvania Railroad, the Radio Corporation of America, Western Union, and Union Switch & Signal in 1946 to show that the control office could be separated from the active territory by great distances and various forms of communications. The transmission path between the office and the active territory in the demonstration included carrier frequencies on railway line wires, Western Union telegraph lines, and several sections of microwave.

The 90-mile CTC installation which was made on the Orinoco Project in Venezuela, with the co-operation of the Radio Corporation of America, in 1953, is novel with respect to all other installations because it uses microwave principles instead of line wires.

A recent and significant development has been made in the control machine to facilitate the work of the dispatcher. Fig. 5 shows this new concept of machine which has the track diagram compressed in the ratio of approximately 4 to 1 with respect to previous machines. This was made possible by removing the push buttons and levers from direct physical relationship on the panel to the functions being controlled. A relatively small push-button panel is located on the desk portion of the machine. By operation of the push buttons on this one panel the dispatcher may issue controls to any desired way station. A logical arrangement in the numbering of way stations and functions permits the operator to issue his controls without the need of reference to a chart or the memorizing of code letters or numbers.

The material which has been presented thus far applies mainly to systems for the handling of traffic on single-track lines of the Western and Southern parts of our country. These same general principles have been applied in the reduction from

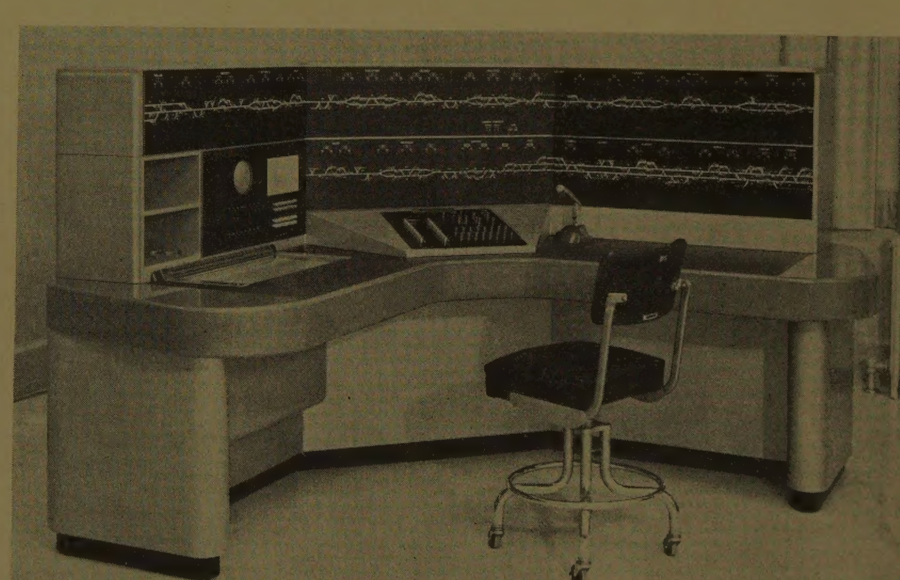


Fig. 5. New concept in traffic control machines

two tracks to a single track on some railroads. In some other instances there are reductions from three and four tracks to two tracks.

When the broad principles of CTC are applied to a railroad with large traffic density, such as the rapid transit lines in New York City, it is necessary to make important changes in the code control and indication system. Because of the relatively short distances involved there is not the same urgency to place many facilities on one pair of wires. There is the requirement for the transmission of more information and in less time. This is accomplished by using systems with greater speed and fewer stations. The latest systems permit controls to the field simultaneously with indications to the office.

This brief reference to the systems on rapid transit lines is made to emphasize the range of application of the basic principle of directing trains by signal indications and to point out that systems have been developed to implement this principle whether it be on a small stretch of track in a remote area, hundres of miles of medium traffic density, or on the dense traffic of rapid transit lines.

In the advancement of these developments there were always definite goals set for the purpose of steering the developments along logical lines. One important goal that is being taken into account at present, and in part guiding development programs, is the crewless train. It is obvious that CTC as discussed in this paper constitutes one of the important components in the operation of crewless trains. The system transmits the desires of the dispatcher to the signal system and transmits information regarding the signals, switches, and trains back to the dispatcher. By the techniques used in cab signaling, the control information from the dispatcher may be applied to the rails and transferred inductively to the cab of the locomotive. Relays on the locomotive controlled in this manner by the dispatcher may have contacts in the locomotive control circuits. Thus the dispatcher would have remote control of the locomotive and safety of operation would be provided by the way side and cab-signaling equipment.

Conclusions

The material presented in this pap and extensive practical experience wit various installations have led to a num ber of conclusions.

1. The direction of trains from a centr control point by control of signal indic tions has advantages over previous meth ods whether the principles be applied t light- or heavy-traffic density railroads, in cluding rapid transit lines.

2. In general a single-track railroad wit this method of control will have approx mately 80% of the traffic capacity of double-track railroad without this metho of control.

3. After CTC has been installed on single-track railroad, the improvement i the movement of a freight train is indicate by a saving of approximately one minute fo each mile of length of the installation.

4. The principles of CTC are being applie in some foreign countries to expand th traffic capacity of existing railroads or ar being included in the construction program of new railroads.

5. From a practical standpoint the tim may be quite near when crewless trains ar operated on some rapid transit lines.

Power Apparatus and Systems—June 1959

Conference Papers Open for Discussion

The conference papers listed below have been accepted for AIEE Transactions and are now open for written discussion until September 29. Duplicate double-spaced typewritten copies of each discussion should be sent to Edward C. Day, Assistant Secretary for Technical Papers, American Institute of Electrical Engineers, 33 West 39th Street, New York 18, N. Y., on or before September 29.

Preprints may be purchased at 40¢ each to members; 80¢ each to nonmembers if accompanied by remittance or coupons. Please order by number and send remittance to:

AIEE Order Department
33 West 39th Street
New York 18, N. Y.

58-89 Application of Switching Transistors and Saturable Reactors in a High-Performance Servo.....Cox, Johannessen

59-505 Deep Ground Beds for Cathodic Protection..Trouard, Wagner, Jr.

AIEE PUBLICATIONS

Electrical Engineering

Official monthly publication containing articles of broad interest, technical papers, and three news sections: Institute Activities, Current Interest, and Electrical Engineering Education. Automatically sent to all members and enrolled students in consideration of payment of dues.

* Subscription price and $1.00 extra for foreign postage both payable in advance in New York exchange.

	Member Prices	Nonmember Prices
Electrical Engineering		annually $12* per year
		Single copies $1.50

Bimonthly Publications

Containing all officially approved technical papers collated with discussion (if any) in three broad fields of subject matter as follows:

	Member Prices	Nonmember Prices
	annually	annually
Communication and Electronics	$5.00†	$8.00‡
Applications and Industry	$5.00†	$8.00‡
Power Apparatus and Systems	$5.00†	$8.00‡

† Members may receive one subscription to any one of the bimonthlies for $2.50. The balance of the $5.00 subscription price shown above will be paid by application of one's annual dues for the year of the subscription. (Members may not reduce the amount of their dues payment by reason of nonsubscription.) Additional subscriptions will be at the $5.00 rate shown above.

‡ Subscription price and 50 cents extra for foreign postage both payable in advance in New York exchange.

	Member Prices	Nonmember Prices
Single copies may be obtained when available.	$1.50 each	$1.50 each

AIEE Transactions

An annual volume in three parts containing all officially approved technical papers with discussions corresponding to six issues of the bimonthly publication of the same name bound in cloth with a stiff cover.

	Member Prices	Nonmember Prices
	annually	annually
Part I Communication and Electronics	$4.00	$ 8.00**
Part II Applications and Industry	$4.00	$ 8.00**
Part III Power Apparatus and Systems	$4.00	$ 8.00**
Annual combination subscription to all three parts (beginning with vol. 77 for 1958).	$10.00	$15.00***
Annual combination subscription to any two parts.		$12.00***

** Subscription price and 75 cents for foreign postage both payable in advance in New York exchange.

*** Subscription price and $1.00 extra for foreign postage both payable in advance in New York exchange.

Electrical Engineering and Transactions

	Member Prices	Nonmember Prices
An annual combination subscription to both publications (effective August 1, 1958).		$24.00§

§ Subscription price and $2.00 extra for foreign postage both payable in advance in New York exchange.

AIEE Standards

Listing of Standards, test codes, and reports with prices furnished on request.

Special Publications

Committee reports on special subjects, bibliographies, surveys, and papers and discussions of some specialized technical conferences, as announced in ELECTRICAL ENGINEERING.

Discount 25% of above nonmember prices to college and public libraries. Publishers and subscription agencies 15% of above nonmember prices. For available discounts on Standards and special publications, obtain price lists from Order Department at Headquarters. Send all orders to:

Order Department
American Institute of Electrical Engineers
33 West 39th Street, New York 18, N. Y.